

Universidade de Lisboa
Faculdade de Ciências
Departamento de Estatística e Investigação Operacional



**A influência do consumo de álcool no insucesso escolar dos
alunos do ensino secundário de São Vicente – Cabo Verde**

Crisolita Sousa de Brito

Dissertação orientada pela
Prof.^a Doutora Maria Fernanda Nunes Diamantino

2012

AGRADECIMENTOS

Os meus agradecimentos vão para todos os que, das mais variadas formas, contribuíram para que esta dissertação chegasse ao resultado final que ora se apresenta.

Um especial agradecimento à Professora Doutora Fernanda Diamantino, orientadora da dissertação, pela disponibilidade manifestada, pela preciosa orientação e pela amizade demonstrada ao longo deste percurso.

Aos meus colegas do curso de mestrado, uma palavra de agradecimento pela solidariedade demonstrada desde a minha atrasada chegada à turma. Uma palavra de agradecimento à coordenadora e ao corpo docente do mestrado em Bioestatística.

Aos directores das escolas secundárias públicas de São Vicente pela autorização cedida para a aplicação do questionário e, aos alunos das mesmas, pela prontidão em responder ao questionário, sem os quais este trabalho não seria possível. Um profundo obrigado!

Agradeço aos meus amigos e colegas de profissão pelas sugestões e pela preciosa ajuda na recolha dos dados.

Um profundo agradecimento aos meus familiares pelo acolhimento, carinho e incentivo ao longo deste percurso.

Finalmente, um especial agradecimento à Suellen – minha filha, minha alegria e meu incentivo - a quem dedico este trabalho.

RESUMO

O álcool é a substância mais consumida pelos adolescentes e jovens, e a idade do início e o padrão de consumo têm sido uma das preocupações, particularmente, dos sectores da saúde e da educação. Este consumo está associado a um conjunto de consequências negativas para a vida do adolescente, entre as quais a dificuldade de aprendizagem e o baixo rendimento escolar. Portanto, a prevenção do consumo de álcool nas escolas, constitui uma via consensual para tentar controlar o problema. Esta dissertação tem como objectivo conhecer o padrão de consumo de álcool dos alunos do ensino secundário em São Vicente (Cabo Verde) e estudar a sua relação com o insucesso escolar. Aplicámos um questionário anónimo a uma amostra de 500 alunos, com idade entre 12 e 21 anos, das 5 escolas secundárias públicas da ilha. Após uma análise exploratória dos dados e uma análise univariada verificou-se que o consumo de álcool podia estar relacionado com o insucesso escolar. Neste sentido, construímos modelos de regressão logística para estudar tal relação. Os resultados evidenciam que o primeiro contacto com o álcool ocorre numa idade precoce e o padrão de consumo, genericamente, varia em função da faixa etária e do género. Entretanto, independentemente do género e da faixa etária, o consumo ocorre fundamentalmente aos fins – de – semana e na companhia dos amigos. Após o controlo dos factores de confusão, o consumo de álcool permaneceu como um factor de risco para o insucesso escolar e verificámos que a probabilidade de um aluno ter insucesso escolar aumenta com a frequência de consumo.

Os resultados do estudo alertam para a necessidade de uma intervenção de toda a comunidade educativa e dos sectores da saúde no sentido de implementar medidas que visem o combate ao consumo de álcool entre os alunos, e adolescentes de uma forma geral.

Palavras-chaves: *Adolescentes, consumo de álcool, insucesso escolar, regressão logística, odds ratio.*

ABSTRACT

Alcohol is the most consumed substance among adolescents and, the onset of consume at an earlier age and the pattern consume is one of the major concerns of the health and education fields. Alcohol use is associated with a difficulties learning and poor academic performance. Starting by the prevention of alcohol consumption at schools is a consensual way of trying to control the problem. The purpose of the present study was to know a pattern of alcohol consumption of the students from São Vicente's high school and its relationship to academic failure. We studied, by an anonymous questionnaire, 500 students from 5 secondary schools aged between 12 and 21. After an exploratory analysis of the sample and a univariate analysis we saw that alcohol could be related to academic failure. In order to study this relationship, we constructed logistic regression models. The results showed that the first intake occurs at the early age and the consumption pattern varies with age and gender. However, consumption occurs mainly at weekends and in company of friends. Alcohol intake is a risk factor of the occurrence of school failure and the odds ratio increases with the consumption's frequency.

This study results apply to the urgency of a intervention by educational community, health sectors and to implement measures aimed to prevent problems related to alcohol consumption among students, and adolescents in general.

Keywords: *adolescents, alcohol intake, academic failure, logistic regression, odds ratio.*

CONTEÚDOS

AGRADECIMENTOS	iii
RESUMO	v
ABSTRACT	vi
CONTEÚDOS	vii
LISTA DE FIGURAS	ix
LISTA DE TABELAS	x
CAPÍTULO I: INTRODUÇÃO	1
1. Objectivos do estudo	3
2. Estrutura da dissertação	4
CAPÍTULO 2: REVISÃO DA LITERATURA	5
1. Breves considerações sobre o insucesso escolar	5
2. O álcool e a adolescência	8
2.1. O consumo de álcool por adolescentes em Cabo Verde	11
3. O álcool, a adolescência e o insucesso escolar	13
CAPÍTULO 3: MODELO DE REGRESSÃO LOGÍSTICA	17
1. Introdução	17
2. Descrição do modelo de regressão logística	21
2.1. Função de ligação	24
2.2. Estimação dos parâmetros – Função de verosimilhança	27
2.3. Teste de significância dos coeficientes	29
2.4. Intervalo de confiança para os parâmetros	32
2.5. Interpretação dos parâmetros	33
2.6. Interacção e factores de confusão	36
2.7. Selecção e validação do modelo	38

2.8. Qualidade do ajustamento do modelo	41
CAPÍTULO 4: METODOLOGIA.....	49
1. Tipo de estudo.....	49
2. População do estudo e técnica de amostragem	49
3. Instrumentos e procedimentos estatísticos	50
4. Variáveis consideradas.....	52
5. Limitações do estudo.....	54
CAPÍTULO 5: APRESENTAÇÃO DOS RESULTADOS.....	57
1. Análise exploratória dos dados	57
1.1. Caracterização do comportamento de consumo de álcool	62
1.2. Descrição do insucesso escolar em função do consumo.....	73
2. Um modelo de regressão logística para análise da associação entre o consumo de álcool e o insucesso escolar	75
2.1. Selecção dos modelos	76
2.2. Comparação dos modelos ajustados.....	82
CAPÍTULO 6: DISCUSSÃO DOS RESULTADOS	87
CAPÍTULO 7: CONCLUSÃO.....	97
BIBLIOGRAFIA.....	101
ANEXOS.....	107

LISTA DE FIGURAS

Figura 1. Variabilidade da idade segundo o género.....	57
Figura 2. Variabilidade da idade segundo o ano de escolaridade	58
Figura 3. Percurso escolar no ensino básico e no ensino secundário	60
Figura 4. Insucesso escolar segundo o género	60
Figura 5. Distribuição da amostra por faixa etária.....	62
Figura 6. Primeiro consumo de álcool segundo o género e a faixa etária.....	63
Figura 7. Local do primeiro consumo segundo o género.....	64
Figura 8. Local do primeiro consumo segundo a faixa etária.....	64
Figura 9. Companhia no primeiro consumo segundo a faixa etária e o género	65
Figura 10. Distribuição do consumo por género e por faixa etária.....	65
Figura 11. Quantidade consumida segundo a faixa etária	67
Figura 12 . Quantidade consumida segundo o género	68
Figura 13. Consumo binge nos últimos 30 dias antes da aplicação do questionário segundo a faixa etária.....	69
Figura 14 . Consumo binge nos últimos 30 dias antes da aplicação do questionário segundo o género.....	69
Figura 15. Curva ROC para cada um dos modelos ajustados	84

LISTA DE TABELAS

Tabela 1. Características de alguns modelos lineares generalizados	20
Tabela 2. Tabela de contingência para testes de diagnóstico	45
Tabela 3 . Representação dos diferentes ciclos e escolas secundárias na amostra	50
Tabela 4. Insucesso escolar segundo alguns atributos	61
Tabela 5. Idade ao primeiro consumo segundo o género e a faixa etária	63
Tabela 6 . Consumo e companhia de consumo segundo o género e a faixa etária	66
Tabela 7. Dias de semana em que ocorre o consumo segundo o género e a faixa etária.....	66
Tabela 8. Frequência de consumo segundo o género e a faixa etária	67
Tabela 9. Frequência de embriaguez segundo o género e a faixa etária.....	70
Tabela 10. Bebida mais consumida segundo o género e a faixa etária.....	70
Tabela 11. Motivações do consumo segundo o género e a faixa etária	71
Tabela 12. Aspectos prejudicados pelo consumo de álcool segundo o género e a faixa etária..	71
Tabela 13. Consumo familiar e consumo dos amigos próximos segundo o consumo, o género e a faixa etária.....	72
Tabela 14. Frequência de ocorrência de algumas consequências do consumo de álcool	73
Tabela 15. Insucesso escolar segundo o consumo, a frequência de consumo e a quantidade consumida	74
Tabela 16. Insucesso escolar segundo a frequência de consumo binge.....	74
Tabela 17. Resultados da análise univariada – relação entre o insucesso escolar e cada uma das covariáveis	76
Tabela 18. Associação entre o consumo de álcool e o insucesso escolar antes e depois da aplicação do propensity score	81
Tabela 19. Medidas de qualidade de ajustamento dos modelos	83
Tabela 20. Odds ratio das covariáveis e seus respectivos intervalos de confiança nos modelos A e C	85

CAPÍTULO I: INTRODUÇÃO

Todo o consumo se inscreve num conjunto de regras de saber viver e de código de civilidade. A fuga a tais regras pode expor a sociedade a situações que prejudicam a sua própria qualidade de vida. Nas últimas décadas o consumo de álcool vem aumentando, constituindo, deste modo, um motivo de preocupação mundial pelos riscos que pode acarretar à saúde e não só. De acordo com a Organização Mundial de Saúde (OMS) a mortalidade e a limitação da condição funcional social associada ao consumo de álcool superam as que estão associadas ao tabagismo, à violência, à SIDA e a outras doenças.

O consumo de álcool espalha-se pelos diversos segmentos sociais, inclusivé a classe dos adolescentes, e a impressão que temos é que na adolescência chega a atingir patamares maiores em relação ao consumo entre adultos.

Em Cabo Verde a situação é preocupante. Apesar de existirem poucos dados estatísticos, é ponto assente que o consumo vem aumentando, e é elevado. Dados datados de 2002 apontaram para 40.3% da população como sendo consumidora de álcool. Segundo os dados do Ministério da Saúde, de 2004 a 2007, houve um aumento significativo de óbitos na população originados pelo consumo de álcool.

Já em 2007, o inquérito sobre as doenças não transmissíveis levado a cabo pelo Instituto Nacional de Estatística refere que 53.0% da população consome álcool, sendo que 3.0% das mulheres e 6.0% dos homens consomem de forma excessiva. Em 2008 este fenómeno foi considerado a terceira causa de morte no país (pelas implicações directas e indirectas).

Preocupante ainda é o facto deste consumo estar a aumentar entre os mais novos, atingindo uma idade cada vez mais baixa. Mesmo não havendo dados estatísticos recentes, o inquérito sobre factores de risco de doenças não transmissíveis realizado nas escolas secundárias de Cabo Verde, em 2005, pela organização não-governamental Associação Zé Moniz (com o apoio da Universidade Nova de Lisboa/ Instituto de Higiene e Medicina Tropical) mostra que 39.3% dos

inquiridos, com idade entre 13 e 16 anos, consumia álcool e 18.0% dos jovens com idade compreendida entre 6 e 11 anos já experimentaram álcool.

Embora a lei cabo-verdiana proíba a venda de bebidas alcoólicas a menores de 18 anos, o consumo de álcool na camada juvenil, e adolescentes em particular, é da consciência nacional e tem sido motivo de preocupação para os sectores da saúde e de diversas componentes da sociedade cabo-verdiana. A Associação de Prevenção do Alcoolismo, a IGAE (Inspeção Geral das Actividades Económicas) e a associação A PONTE, ao lado de outros sectores, têm agido na luta contra tal flagelo com campanhas e palestras, entre outras actividades. Entretanto, paradoxalmente, em oposição às acções de prevenção e às denúncias dos efeitos maléficos do consumo de álcool, existe um apelo consagrado à sua comercialização e, até certo ponto, ao seu consumo. A acessibilidade, a indústria da propaganda e a própria cultura da sociedade têm incrementado o consumo de álcool, tornando este mais acessível à camada jovem, e aos adolescentes em particular.

A adolescência, caracterizada por consideráveis transformações, é também uma fase de passagem à vida adulta. Esta fase de preparação para a vida adulta não se limita apenas à aquisição de conhecimentos, mas também de valores e hábitos que culminam com uma boa integração na sociedade e uma preservação da saúde, entre outros aspectos. Por outro lado, dadas as particularidades da adolescência, entre as quais a maturidade em fase de concretização, os adolescentes são vulneráveis a comportamentos de risco e, portanto, cabe aos adultos, em particular aos serviços da educação, da saúde e a própria família, interferirem de modo a minimizar qualquer desvio de comportamentos e atitudes saudáveis. O consumo precoce de álcool é uma das realidades desta fase que põe em risco esta concretização.

Para além de outros danos à saúde, o efeito do álcool no cérebro nesta faixa etária é muito mais devastador se compararmos com o efeito no cérebro do adulto, pois atinge regiões ligadas à memória, à aprendizagem, ao autocontrole e à motivação. O prejuízo na capacidade de processar novas informações, juntamente com as alterações na capacidade de concentração e retenção, pode prejudicar o desempenho escolar e o desenvolvimento das várias habilidades de muitos adolescentes que fazem uso do álcool.

A escola, que também vivencia o problema do consumo de álcool entre os seus, é um cenário privilegiado para a educação para a saúde e para o saber viver. O êxito escolar está associado ao

bem-estar do aluno que, por sua vez, depende da boa atitude deste. Portanto, a escola actual que assenta nos quatro pilares da educação de Jaques Delors - *aprender a conhecer, aprender a fazer, aprender a conviver com os outros e aprender a ser* - pretende uma compreensão dos factores que influenciam o insucesso escolar no contexto global dos seus educandos.

De acordo com os actuais dados do Ministério da Educação e Desporto de Cabo Verde, as taxas de reprovação no ensino secundário atingiram os 20.1%. O Director Geral do Planeamento, Orçamento e Gestão aponta como causas das retenções um conjunto de variáveis que inclui o comportamento do aluno e a sua capacidade cognitiva, entre outras.

Consideramos complexa e grave a questão do consumo de álcool pelos adolescentes logo, pensamos, enquanto profissionais do sector da educação, que é necessário conhecer a real situação, reflectir sobre o seu efeito no rendimento escolar e tomar importantes acções para informar e consciencializar os adolescentes e a sociedade em geral sobre a questão, bem como delinear estratégias no âmbito da promoção para a saúde. Esta dissertação visa contribuir neste sentido.

Este trabalho é também uma ambição pessoal, como profissional no ramo da Matemática, por ser um exemplo de aplicação da Matemática, da Estatística em particular, como uma ferramenta útil para outros ramos do conhecimento (neste caso específico, da Saúde e da Educação). Neste trabalho usámos o modelo de regressão logística no estudo da relação entre o consumo de álcool e o insucesso escolar. Este modelo tem sido uma das ferramentas estatísticas muito utilizada nas Ciências Sociais e nas Ciências da Saúde.

Atendendo à disparidade geográfica, limitámos o estudo a uma das ilhas do país. Estudámos a influência do consumo de álcool no insucesso escolar através de uma amostra representativa seleccionada entre os alunos das escolas secundárias de São Vicente – Cabo Verde.

1. Objectivos do estudo

Julgamos ser necessário conhecer-se o padrão de consumo de álcool por parte dos nossos alunos e estudar os possíveis efeitos do mesmo de modo a delinear programas de intervenção destinados à prevenção de problemas ligados ao consumo de bebidas alcoólicas bem como à promoção de saúde e sucesso dos alunos. Assim, destacamos como objectivo geral – verificar se o consumo de álcool é relevante na explicação do insucesso escolar dos alunos do ensino

secundário de São Vicente – Cabo Verde. Para a realização do estudo delineámos, também, os seguintes objectivos específicos:

- Estudar a prevalência do consumo de álcool com base numa amostra representativa dos alunos do ensino secundário (público) de São Vicente;
- Estudar a distribuição do consumo por sexo e por faixa etária;
- Estudar a distribuição do consumo por percurso escolar dos alunos;
- Comparar os alunos, por género e por faixa etária, em termos de:
 - Idade, local e companhia no primeiro consumo;
 - Frequência de consumo, quantidade habitualmente consumida e companhia no consumo frequente;
 - Bebida consumida com maior frequência;
 - Frequência de episódios de embriaguez e de consumo *binge*;
 - Motivações do consumo;
 - Frequência de ocorrência das consequências do consumo;
 - Consumo entre os amigos mais próximos;
 - Consumo regular e problemático na família;
- Analisar o efeito do consumo de álcool no insucesso escolar dos alunos ao lado de algumas variáveis associadas ao insucesso escolar.

2. Estrutura da dissertação

O trabalho está estruturado da seguinte forma: o primeiro capítulo dedica-se a uma introdução; no segundo capítulo apresentamos uma revisão da literatura relativamente ao insucesso escolar, ao consumo de álcool na adolescência e seus feitos na aprendizagem, de modo a ajudar na interpretação dos nossos resultados; o terceiro capítulo destina-se a uma descrição do modelo de regressão logística; no quarto capítulo apresentamos os métodos, as variáveis, os procedimentos e as técnicas utilizadas no estudo; no quinto capítulo apresentamos os resultados, descrevendo o comportamento da amostra face às bebidas alcoólicas e faz-se uma aplicação do modelo de regressão logística no estudo da relação entre o consumo de álcool e o insucesso escolar; no sexto capítulo apresentamos a discussão dos resultados; o sétimo (e último capítulo) é dedicado à conclusão do trabalho. No fim, apresentamos a bibliografia e os anexos.

CAPÍTULO 2: REVISÃO DA LITERATURA

1. Breves considerações sobre o insucesso escolar

O insucesso escolar tem sido, ao longo dos tempos, uma preocupação quer política quer para a comunidade educativa, de uma forma geral. A palavra insucesso, de origem latina *insucessu(m)*, significa “mau resultado” ou “mau êxito”. Tais definições levam a encarar o insucesso escolar como um problema do aluno que depende do seu grau de inteligência. Entretanto, com o avanço das investigações na área da educação, viu-se que não é possível estabelecer uma relação de casualidade entre insucesso escolar e inteligência do aluno, e o insucesso escolar deixa de ser um problema da responsabilidade exclusiva do aluno passando a ser um fenómeno social e de carácter massivo (Benavente, 1990).

Geralmente, o insucesso escolar advém do regime anual de aprovação/reprovação dos alunos (atingir ou não metas educativas dentro de um determinado período de tempo estabelecido) que se baseia numa estrutura de avaliação característica do sistema de ensino. Portanto, o insucesso institucionalmente considerado baseia-se numa taxa de reprovação (ou mesmo de abandono escolar). Entretanto, o insucesso escolar vai além do certificado pela escola pois, e segundo Formosinho (1991), a educação tem como objectivo não só transmitir conhecimentos e técnicas como também transmitir normas, valores, crenças, hábitos e atitudes e promover o desenvolvimento integral do educando. Assim, e segundo o referido autor, o insucesso educativo individual pode ser na instrução, na socialização ou ainda na integração.

Outros autores também defendem que o sucesso/insucesso escolar não pode ser medido apenas pelo *feedback* académico, mas também pelo seu desenvolvimento integral, pessoal e social, pela maturidade e pela capacidade de estabelecer relações humanas entre outros aspectos, bem como pelas relações que a escola estabelece com os diversos alunos e com a dificuldade que a escola tem em se relacionar com os que são social e culturalmente diferentes (Benavente, 1990; Silva, 2007 e Tomé & Matos, 2006, citados por Alves, 2010). Assim, se inicialmente a preguiça,

a falta de capacidade ou a falta de interesse do aluno eram as únicas causas do insucesso escolar apontadas, hoje a este conjunto adicionam-se os programas de ensino, o currículo escolar, as metodologias e as respectivas estratégias utilizadas, as matérias e recursos adoptados, os professores, os equipamentos escolares, as modalidades gerais do sistema de avaliação, a família e a origem social, o estilo de vida dos alunos e a própria sociedade (Borone, 2001; Bissoto, 2009 e Silva, 2007, citados por Alves, 2010).

Nesta óptica, entende-se que o insucesso escolar é complexo e, conseqüentemente, não é fácil de se medir, podendo cobrir realidades diversas e ser condicionado por contextos históricos específicos. Uma vez que as razões do insucesso escolar são várias e estas não estão isoladas umas das outras, como sugerem Facci *et al.* (2007), para melhor entendermos o insucesso escolar não devemos restringir-nos aos limites do quotidiano da escola, mas sim atender aos problemas da comunidade no sentido de compreender na totalidade os factores associados ao insucesso.

Como já referido, um dos factores que pode influenciar o sucesso/insucesso escolar é a qualidade individual do educando. Entretanto, o insucesso escolar também pode ter algum efeito emocional no aluno (como, por exemplo, na sua auto-estima ou na motivação para seguir adiante com os estudos). Quanto maior é a dimensão do problema emocional (e não só) do aluno, maior é a dificuldade que este terá em responder positivamente às suas tarefas, em particular, às académicas. Assim, seja o insucesso causa de um distúrbio emocional ou consequência de alguma limitação ou da realidade própria do aluno, resolver o problema do aluno poderá implicar, de uma certa forma, um combate ao insucesso escolar. Nesta óptica, não podemos distanciar a questão do insucesso/sucesso escolar de um aspecto de saúde (mesmo não o considerando uma doença).

Segundo a Organização Mundial de Saúde (OMS), os determinantes sociais de saúde são as condições em que as pessoas nascem, crescem, vivem, trabalham e envelhecem, incluindo o sistema de saúde. Mesmo no tocante à própria doença, hoje fala-se num modelo multicausal, segundo o qual as causas biológicas e físicas da doença estão relacionadas com várias outras, incluindo factores ambientais e sociais. Portanto, nenhum aspecto (psicológico, social ou mesmo comportamental) do indivíduo, bem como as condições sociais que o influenciam, podem ser ignorados.

A saúde é, segundo a OMS, “um estado dinâmico de completo bem-estar físico, mental, espiritual e social e não meramente ausência de doença”. Assim sendo, um aluno com insucesso escolar não estará em conformidade com a própria definição de saúde. Mesmo que o aluno esteja (ou acha que está) em conformidade física e psicológica, tendo insucesso escolar já garante a inconformidade com o bem-estar social, uma vez que a norma social preconiza o desempenho escolar como correcto. Portanto, combater o insucesso escolar dos alunos também é prevenção para o seu bem-estar nas componentes da definição de saúde.

Desta forma a educação e a saúde estão relacionadas de tal forma que a frase “*para ter educação é preciso ter saúde e para ter saúde é preciso ter educação*” encontra-se, de certo modo, no interior das práticas escolares (Spazziani, 2001). Entretanto, e segundo Bagnato (1987) (citado por Spazziani, 2001), geralmente a contribuição dos programas para saúde são reduzidos em relação ao que propõe o programa curricular, predominando métodos científicos e acções pedagógicas que não influenciam hábitos e atitudes, embora a saúde escolar tenha como objectivo garantir objectivos necessários para a aprendizagem, reduzindo a taxa de insucesso escolar ou retenções.

Não obstante a diversidade de definições e as várias causas, a reprovação ou retenção do aluno num determinado nível escolar é um dos indicadores mais preocupantes do insucesso escolar. Assim, neste trabalho a retenção do aluno num determinado ano de escolaridade é considerado como insucesso escolar no ano escolar em questão.

A *qualidade de ensino* continua sendo uma das palavras-chave da Educação em Cabo Verde, que optou, para o ano lectivo 2011/2012, o lema “*Juntos para uma educação de qualidade*”. Um dos objectivos da educação é diminuir o insucesso escolar, sobretudo no ensino secundário. Segundo dados nacionais, mais de metade dos alunos do secundário estão na classe errada para a sua idade (ou seja, estão atrasados na escola). Por exemplo, os dados oficiais do governo do ano lectivo 2009/2010 mostram que 78.3% dos alunos no 12º ano já deveriam estar aptos para entrar no sistema universitário. Esta distorção idade-classe está relacionada com as retenções ao longo do percurso escolar (UNICEF & ICCA, 2011).

Portanto, para melhor combater qualquer taxa de insucesso escolar ou para melhorar a qualidade de ensino, pensamos ser importante, e como defende Facci *et al.* (2007), atender à realidade dos nossos educandos que, hoje, estão inseridos numa sociedade que lhes proporciona

um contexto sociocultural e outros factores que podem, de uma forma ou de outra, interferir na sua vida académica.

2. O álcool e a adolescência

O álcool é uma das poucas drogas psicotrópicas que tem consumo livremente aceite na sociedade. Mas, o seu mau uso e abuso têm como consequências o distúrbio biológico, psicológico e social com enorme impacto na sociedade (Aragão & Sacadura, 2002). Hoje, a preocupação assenta no elevado consumo, sobretudo na adolescência e na juventude, que tem as mais variadas causas. De acordo com dados estatísticos da OMS, o álcool causa anualmente 2.5 milhões de mortes (4% de todas as mortes) e, em particular, 9.0% das mortes de jovens com idade entre 15 e 29 anos são atribuídas ao álcool.

A adolescência é uma fase de desenvolvimento físico e emocional, em que o indivíduo adopta comportamentos influenciados pelo meio socio-ambiental e, por isso, é também uma fase de risco para ingressar no consumo das substâncias proactivas, como o álcool. Este hábito de consumo, que normalmente tem o seu início numa actividade associada ao lazer é susceptível de se tornar num hábito de consumo perigoso (pela dependência que pode criar), tornando-se, assim, uma grande ameaça à saúde dos adolescentes.

O álcool é uma substância de fácil acesso e o seu consumo está intimamente ligado a actividades agradáveis, festivas ou de outra natureza social (Freyssinet – Dominjon & Wagner, 2006), o que se reflecte num consumo precoce e em excesso. Estudos revelam que a fase de experimentação do álcool ocorre, normalmente, entre 11 e 12 anos, e apesar de, muitas vezes, ser apenas experimental, pode-se notar padrões de consumo que são observados na vida adulta e que, consequentemente, indicam a necessidade de se adoptar medidas preventivas nesta fase de desenvolvimento.

Num estudo realizado por Alves (2010), com estudantes de 15 a 20 anos, 67.5% dos inquiridos teve a primeira experiência com o álcool numa idade entre 11-15 anos e uma outra percentagem (11%) teve a mesma experiência numa idade ainda mais precoce (5-10anos). Zeigher *et al.* (2005) referem que o consumo de álcool por menores afecta toda a comunidade dos Estados Unidos da América e que os jovens consomem a sua primeira bebida alcoólica aos 12 anos de

idade. Outro estudo (Ferigolo *et al.*, 2004) realizado em Porto Alegre com 402 jovens, de idade entre 10 e 20 anos, revelou que o início do consumo de álcool ocorre, em média, aos 11 anos.

Para além do início precoce de consumo, a literatura revela alguma preocupação relativamente aos padrões de consumo de álcool adoptado pelos adolescentes. Segundo Harnett *et al.* (2000) existem 8 estilos de consumo de álcool: 1. “Chilhood”, 2. “Adolescent”, 3. “Experimental”, 4. “Sociable”, 5. “Recreational”, 6. “Safe”, 7. “Therapeutic” e 8. “Structural”.

O primeiro refere-se ao período em que o consumo de bebidas alcoólicas é controlado pelos pais, quando a primeira experiência com o álcool acontece em casa, durante uma refeição e num momento festivo. O segundo estilo caracteriza-se por ocasiões em que os indivíduos organizam as suas actividades de modo a estarem com os amigos e na ausência dos familiares. Neste estilo, o que conta é a experiência como o todo, passando a ser normal qualquer efeito adverso do álcool (como a ressaca, entre outros). O estilo “Experimental”, que está associado à maioridade e à possibilidade de frequentar as discotecas, caracteriza-se pela procura, por parte do consumidor, de novas experiências com o álcool que são proporcionadas quer por novas bebidas, quer por novos contextos ou mesmo por novos parceiros de consumo.

Os próximos três estilos (“Sociable”, “Recreational” e “Safe”) estão relacionados com um maior controlo por parte do consumidor, o que é evidenciado, segundo os autores, pela transição da adolescência para a fase adulta. No estilo “Sociable”, o consumidor vê o consumo como um facilitador de relações sociais e, portanto, o consumo é moderado. No estilo “Recreational” (em que, geralmente, ocorre a mistura de bebidas), o consumidor desperta o gosto pela embriaguez e planeia encontros com os amigos (geralmente aos fins de semana) para se embriagarem. No estilo “Safe”, o indivíduo pretende conservar o seu bem-estar físico e social, procurando minimizar os riscos de causar danos tanto a si como aos outros. Alguns consumidores diminuem o consumo para manter a lucidez e a eficiência mental e outros podem até consumir quantidades em excesso, porém planeiam a forma de evitar qualquer perigo (como o de conduzir sobre o efeito do álcool).

No estilo “Therapeutic”, o consumo (geralmente solitário) tem por objectivo aliviar algum tipo de mal-estar ou sintoma (como por exemplo, beber para esquecer os problemas, para vencer a timidez e para combater a insónia, entre outros). No último estilo (“Structural”), o consumo de álcool está relacionado com outros factores estruturantes da vida do consumidor (como, por

exemplo, o emprego e o estudo) e o objectivo do consumo é esquecer, mesmo que seja por um curto período de tempo, estes factores.

Genericamente, há informações consistentes sobre os factores que mais influenciam o início ou mantém o consumo de álcool na adolescência, demonstrando que o consumo de álcool entre os adolescentes adopta os diferentes estilos defendido por Harnett *et al.* (2000). A literatura sugere que a pressão dos amigos ou grupo de pares que consomem álcool, a estrutura e o ambiente familiar, a publicidade, a excessiva tolerância dos pais, a própria cultura da sociedade e o fácil acesso e baixo custo do álcool são os factores que mais têm contribuído para o consumo de álcool entre os adolescentes (Carvalho *et al.*, 2007; Peuker *et al.*, 2006; Breda, 1996). Entretanto, não se pode ignorar as características do próprio indivíduo e da própria fase da adolescência (Moreira, 2002).

O fácil acesso do álcool, a sua presença nas festas e diversões e a pressão exercida pelo grupo de pares são alguns factores que fazem com que os adolescentes bebam de forma muitas vezes exagerada (Pillon & Corradi – Webster, 2006). O consumo em quantidades elevadas e num curto espaço de tempo - consumo *binge*- tem sido umas das preocupações reveladas por diversos estudos. Alguns autores (Nelson, Naimi, Brewee & Wechsler, 2005, citados por Alves, 2010) definem o consumo *binge* como sendo o consumo de cinco ou mais bebidas alcoólicas por ocasião, tratando-se de um homem, ou o consumo de quatro ou mais bebidas alcoólicas, caso se trate de uma mulher. Já Marinho (2005), define o consumo *binge* como sendo o consumo de quatro ou mais bebidas alcoólicas tratando-se de rapaz ou de três ou mais bebidas, para o caso das raparigas. No nosso estudo, atendendo à idade dos nossos inquiridos, considerámos como consumo *binge* o defendido por Marinho (2005).

O padrão de consumo *binge* aumenta o risco de o indivíduo desenvolver a dependência e vir a sofrer consequências negativas associadas ao consumo de álcool (Peuker, Fogaça & Bizarro, 2006). Zeigler *et al.* (2005) referem que os indivíduos, com idade entre 18 e 24 anos, que consomem de forma *binge* pelo menos uma vez por semana podem ter, para além dos problemas ligados à saúde, problemas em alcançar objectivos específicos da transição da adolescência para a idade adulta, como o sucesso escolar e o emprego. Assim, e segundo os referidos autores, quando mais cedo se começar o padrão *binge*, maiores são os riscos de problemas na adolescência e no futuro.

O consumo de álcool na adolescência está associado a um conjunto de prejuízos no desenvolvimento do adolescente e em resultados posteriores. Atendendo às questões neuroquímicas da fase de amadurecimento cerebral e às especificidades desta etapa da vida, os prejuízos decorrentes do consumo de álcool nos adolescentes são diferentes dos evidenciados na fase adulta. Um episódio de consumo exagerado de álcool pode causar um “*blockout*” ou perda de memória para acontecimentos que ocorreram durante o episódio de bebida. A perda de memória é, geralmente, temporária mas pode persistir durante algum tempo depois do episódio de bebida que a causou. Geralmente, entre 6 -24 horas após o consumo exagerado ou prolongado de álcool, o consumidor pode sofrer dos sintomas da ressaca, como dores de cabeça, tontura, náusea e vômito, fraqueza dos músculos ou dor, depressão e irritabilidade entre outros. Os processos de pensamento e aprendizagem também ficam afectados (Ziegler *et al.*, 2005). Por outro lado, o consumo de álcool nesta fase da vida aumenta a possibilidade de envolvimento em actividades sexuais sem protecção e, consequentemente, com maior exposição às doenças sexualmente transmissíveis.

2.1. O consumo de álcool por adolescentes em Cabo Verde

A somar às outras bebidas alcoólicas, uma das produções nacionais (particularmente nas duas maiores ilhas do país - Santo Antão e Santiago) é a aguardente (*grogue*), álcool muito comercializado e consumido internamente (sendo, também, um dos produtos exportados do país). Devido ao seu baixo custo, este produto é de fácil acesso para qualquer classe social do país.

Apesar de existirem poucos dados estatísticos, é ponto assente que o consumo de álcool é elevado no país. Se em 2008 o álcool foi considerado a sétima causa de morte em Cabo Verde, dados actuais do Ministério da Saúde apontam-no como a terceira causa (pelas consequências directas e indirectas do seu consumo). Por exemplo, segundo os dados do Hospital da Ribeira Grande - Santo Antão, em 2009 os internamentos, no serviço de medicina, devido ao consumo de álcool ocuparam 20% do total de 683 internamentos, o que retrata a problemática que tem sido o alcoolismo em Ribeira Grande.

As notícias nacionais (pelos diversos meios de comunicação), as actividades levadas a cabo pelas diversas entidades ligadas à problemática do consumo de álcool bem como as preocupações das entidades governamentais dadas a conhecer ao público nacional, mostram

que é assente que o consumo de álcool é elevado e este atinge, também, os mais novos (jovens e adolescentes).

Sendo quase um hábito social e uma prática cultural, o combate ao consumo de álcool não se tem mostrado fácil. Embora seja proibida a venda de bebidas a menores de 18 anos, os jovens são introduzidos neste tipo de consumo quase como um ritual de passagem para a idade adulta e poucos são os que resistem à pressão (des várias ordens). Este consumo, principalmente na faixa etária mais nova, está associado às diversões e saídas nocturnas.

Apesar dos reiterados alertas, o consumo abusivo de álcool entre os cabo-verdianos é cada vez mais preocupante tanto pelo padrão como pela precocidade do início de consumo. Num estudo nacional datado de 2005, realizado pela organização não-governamental Associação Zé Moniz, 39.3% dos inquiridos com idade entre 11 e 13 anos referiram consumir álcool e 18% dos inquiridos com idade entre 6 e 11 anos referiram já ter experimentado álcool. Outro estudo recentemente realizado numa das escolas secundárias da ilha de São Vicente, com alunos entre 14 e 21 anos de idade, revelou que metade dos inquiridos consome álcool, e 18.0% iniciaram o consumo de álcool antes dos 15 anos, sendo que a idade mínima do início de consumo foi de 9 anos (Antunes e Caçador, 2012). Ainda segundo o mesmo estudo, 40.0% dos alunos consome álcool uma vez ou mais por semana.

No sentido de enfrentar e de combater esta ameaça à saúde pública, algumas organizações nacionais têm realizado actividades alusivas ao consumo de álcool. Por exemplo, em Novembro de 2011, a Associação Cabo-verdiana de Prevenção de Alcoolismo (ACPA) realizou um debate alusivo ao tema “Alcoolismo: uma urgência nacional”, no qual o presidente nacional deixou o seguinte alerta:

"O sofrimento que o alcoolismo vem causando num número apreciável de famílias, as grandes repercussões ao nível do trabalho, da saúde das pessoas e do desenvolvimento da juventude, fazem com que ele seja considerado, de modo apropriado, uma urgência nacional (...) Este esforço deverá abranger as áreas da educação, da prevenção, da legislação e regulamentação, do controlo da produção, dos preços das bebidas alcoólicas, da publicidade, bem como do tratamento e da reinserção social."

(Jorge Carlos Fonseca, Presidente de Cabo Verde)

Expresso das Ilhas, 11.11.2011.

Ao lado da forte promoção do consumo de álcool no país (particularmente, através da publicidade e dos eventos culturais), outras associações e sectores da saúde têm levado a cabo actividades diversas de sensibilização para atitudes saudáveis referentes ao consumo de álcool. Entretanto, o consumo de álcool não tem demonstrado sinais de diminuição, o que nos leva a pensar que é urgente e necessário um trabalho mais profundo no sentido de combater o elevado consumo, em particular, o consumo entre os jovens e adolescentes. Acreditamos que o sector da educação deve ser um veículo para tal propósito, sendo imprescindível uma actuação em parceria dos sectores da educação, da saúde e da legislação, bem como uma participação dos pais/encarregados de educação.

Este trabalho surge, portanto, no intuito de oferecer uma base que possa ser útil para dar a conhecer a realidade do consumo de álcool entre os alunos e sugerir a referida parceria, capaz de levar a cabo estratégias para o combate ao consumo de álcool entre os nossos alunos e, consequentemente, aos efeitos que este tem para a vida do adolescente, em particular a saúde e o rendimento escolar.

3. O álcool, a adolescência e o insucesso escolar

A adolescência vai dos 12 aos 20 anos de idade (Harper & Marshel, 1991) e é marcada, entre outros aspectos, pelo contínuo desenvolvimento do hipocampo (estrutura localizada nos lobos temporais do cérebro humano, considerada a principal área da memória e importante componente do sistema límbico, responsável pelas emoções). Segundo Graf (2004) (citado por Alves, 2010), o desenvolvimento do cérebro só está efectivamente concluído por volta dos 20 anos e, portanto, antes dessa idade os indivíduos estão mais expostos aos malefícios do consumo de álcool, uma vez que o álcool diminui a sensibilidade dos centros nervosos e, consequentemente, interfere na capacidade de pensar, no poder de memorizar e de manter a atenção.

Assim, a ingestão contínua e persistente de álcool é particularmente prejudicial na fase da adolescência (Peuker *et al.*, 2006) podendo provocar alterações significativas ao nível da aprendizagem e da memória (Ribeiro, 2008). Alves e Kossobudzy (2002) referem, também, que o consumo de álcool altera as funções cognitivas e de memória, as formas de pensamento e percepções, o que influencia a aprendizagem.

Segundo uma investigação realizada por Ryback (1971), a magnitude dos efeitos do álcool na memória aumentam com a quantidade consumida. Quando as doses são em pequenas quantidades ou moderadas (tais como as que produzem concentrações de álcool no sangue abaixo de 0.15%) as debilidades de memória tendem a ser menores e moderadas. À medida que as doses aumentam, as debilidades de memória resultantes da ingestão de álcool podem tornar-se mais profundas, muitas vezes culminando com a perda de consciência e a incapacidade para recordar elementos críticos de acontecimentos (White *et al.*, 2002, citado por Zeigher *et al.*, 2005).

Sabendo que a memória é a função fundamental no processo de aprendizagem e que esta se altera com o consumo de álcool, é natural que o álcool comprometa o processo de aprendizagem (Pechansky *et al.*, 2004) levando ao insucesso escolar. Por outro lado, e segundo Zeigler *et al.* (2005), o álcool pode perturbar o ciclo sono-vigília, alterando o tempo total do sono, o tempo necessário para adormecer e a sequência e a duração dos estados do sono e, consequentemente, pode provocar a sonolência durante o dia e limitar o desempenho escolar. O insucesso escolar é mais do que uma retenção num determinado ano lectivo podendo, também, verificar-se quando as classificações não correspondem às expectativas do aluno ou quando o rendimento escolar é inferior ao desejado (ou mesmo inferior ao que o aluno é capaz). Assim, os adolescentes que consomem álcool poderão ver comprometido o seu desempenho escolar, devido às faltas às aulas ou ao fraco rendimento nas aulas e nas avaliações, muitas vezes consequências do aluno ter chegado atrasado às aulas ou mesmo por dormir nas aulas porque consumiu na noite (ou no dia) anterior (Pillon & Corradi - Webster, 2006). Consequentemente, o risco de reprovação aumenta conforme aumenta a frequência e a quantidade consumida (Aertgeerts & Buntix, 2002, citados por Pillon & Corradi - Webster, 2006).

No estudo realizado entre jovens oriundos dos distritos de Vida Real e de Bragança, Alves (2010) concluiu que um aluno que consome álcool apresenta uma probabilidade duas vezes maior de ter retenções do que um abstémico e, entre os que consomem álcool, um aluno que apresenta o padrão de consumo *binge* tem uma probabilidade duas vezes maior de ficar retido no seu percurso escolar quando comparado com aquele que não apresenta o padrão *binge*.

Segundo estudos (Webb *et al.*, 2007), a relação entre o consumo de álcool e o baixo rendimento escolar está mais relacionada com os jovens do sexo masculino, relativamente aos do sexo feminino. Na opinião dos autores, tal diferença deve-se tanto ao facto de os rapazes ingerirem

maiores quantidades de álcool como ao facto de a ingestão de álcool ser mais frequente quando estes são comparados com as raparigas. Já no estudo realizado por Mendes e Lopes (2007), apesar de os autores não encontrarem uma associação significativa entre o consumo de álcool e o rendimento escolar dos 272 jovens inquiridos, os alunos sem insucesso escolar apresentaram maior taxa de abstinência face às bebidas alcoólicas quando comparados com os que apresentam retenções. Ainda no mesmo estudo, os alunos com retenções apresentaram uma maior percentagem na categoria de “consumidor habitual”.

A literatura sugere que o consumo de álcool está, também, associado a um conjunto de outros problemas escolares - a indisciplina, o abandono escolar, o absentismo às actividades lectivas e a fraca dedicação às actividades escolares, entre outros (Silva *et al.*, 2008; Anguita & Lizana, 1996). Segundo Gaspar e Matos (2008), os adolescentes que nunca ou raramente consumiram álcool são os que mais referem gostar da escola. Por outro lado, os adolescentes mais satisfeitos com a escola revelam menos probabilidade de desenvolver comportamentos de risco, como o consumo de álcool.

Alguns autores partilham a opinião de que o insucesso escolar pode também predizer o início do consumo de álcool (ou de outras drogas). O absentismo, a falta de motivação para os estudos, o fraco envolvimento escolar, a má adaptação escolar que têm efeitos no rendimento escolar, são factores que podem predispor os adolescentes a consumir drogas, entre os quais o álcool (Gaspar & Matos, 2008). Por outro lado, quando o insucesso escolar é atribuído à falta de capacidade, tal pode afectar a auto-estima do aluno, levá-lo ao abandono escolar e, consequentemente, a maior probabilidade de entrar no mundo do consumo de álcool.

Existindo uma relação entre o insucesso escolar e o consumo de álcool por parte dos adolescentes, e independentemente de qual fenómeno desencadeia o outro, pensamos ser necessária a actuação dos diferentes sectores da sociedade, entre os quais a escola, no sentido de conhecer a real situação de consumo e agir com estratégias de combate e prevenção de modo que o combate ou a diminuição de um destes fenómenos (consumo de álcool ou insucesso escolar) seja, consequentemente, o combate do outro fenómeno e a prevenção para a saúde dos adolescentes.

CAPÍTULO 3: MODELO DE REGRESSÃO LOGÍSTICA

1. Introdução

Nos estudos estatísticos, cujo objectivo é estudar uma relação entre variáveis, encontramos dois tipos de variáveis, conforme o papel que desempenham: a *variável resposta* ou *dependente* - que representa a variável de interesse - e as *variáveis explicativas* ou *independentes* (denominadas também de *covariáveis*). As variáveis (dependentes ou independentes) podem ser contínuas, discretas ou dicotómicas, e as independentes podem ser, ainda, ordinais.

Estabelece-se uma associação entre duas variáveis quando se verifica que os valores de uma são diferentes em função dos valores da outra. Por vezes interessa-nos conhecer a variabilidade da variável resposta através de uma covariável ou da combinação linear dos valores de um conjunto de covariáveis. Os modelos de regressão constituem uma das ferramentas estatísticas que nos permitem estudar tal variabilidade.

Um *modelo* é uma descrição do tipo de relação particular entre diferentes variáveis. Na realidade, um modelo não descreve exactamente a realidade, mas sim, aproximadamente. Portanto, é importante avaliar, para cada situação, o modelo mais simples que melhor descreve a realidade. No caso de existir mais do que uma variável explicativa, são úteis os *modelos multivariados* que têm uma enorme aplicação em investigações de diversas áreas pois permitem avaliar simultaneamente o efeito das várias variáveis explicativas sobre a variável resposta. Neste caso, pode-se entender que a influência de cada variável explicativa é “controlada” pelo efeito das demais.

Antes de avançar, pensamos ser útil definir algumas notações e terminologias a utilizar ao longo do trabalho. A variável resposta é uma variável aleatória Y e o conjunto das k variáveis independentes é definida pelo vector $\mathbf{x} = (x_1, x_2, \dots, x_k)$. Assumimos que os dados estão estruturados da forma (Y, \mathbf{x}) em que o par (y_i, \mathbf{x}_i) , com $i=1, \dots, n$, representa o conjunto das

observações referente ao i -ésimo indivíduo. Admite-se ainda que as n componentes do vector $\mathbf{Y}=(Y_1, Y_2, \dots, Y_n)^T$ são independentes.

Um modelo de regressão linear que relaciona Y com as k variáveis independentes é dado pela equação matemática:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \xi \quad (3.1)$$

em que os $(k+1)$ parâmetros β_j são os coeficientes de regressão (que quantificam a variação de x_j sobre a variável dependente) e ξ é o erro aleatório (diferença entre o valor observado e o valor ajustado pelo modelo). Este modelo é designado *modelo de regressão linear multivariada*. Em particular, quando $k=1$ o modelo é designado *modelo de regressão simples* ou *univariada*.

Numa regressão pretende-se particularmente (se não, principalmente) conhecer o valor médio da variável resposta, dados os valores das covariáveis. Este valor médio condicional é expresso por $E(Y | \mathbf{x})$ e pode ser definido em função de \mathbf{x} :

$$E(Y | \mathbf{x}) = \beta_0 + \sum_{j=1}^k \beta_j x_j.$$

A aplicação do modelo de regressão dado em (3.1) pressupõe a verificação de alguns pressupostos que podemos sintetizar em:

- A variável resposta Y segue uma distribuição normal;
- As variáveis explicativas x_1, x_2, \dots, x_n não são correlacionadas (ausência de multicolinearidade entre as variáveis explicativas);
- Os erros aleatórios são independentes;
- ξ segue uma distribuição normal de valor médio nulo e variância σ^2 .

Várias são as situações em que o fenómeno em estudo não apresenta uma variável resposta para a qual seja possível a suposição de normalidade, ou ainda em que os erros não apresentam uma distribuição normal. Nestes casos, é necessário uma transformação da variável resposta de

modo a obter a normalidade pretendida. Uma das propostas sugeridas foi apresentada por Nelder e Wedderburn (1972) que propuseram os *modelos lineares generalizados* (MLG).

O modelo linear generalizado é definido por uma variável dependente cuja distribuição de probabilidade é membro da *família exponencial*, um conjunto de variáveis independentes descrevendo a estrutura linear do modelo e uma função de ligação entre a média da variável dependente e a estrutura linear.

Seja Y_i uma variável aleatória, Y_i pertence à *família exponencial* se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) se escrever na forma:

$$f(y_i | \theta_i, \phi_i) = \exp \left\{ \frac{y_i \theta - b(\theta_i)}{a(\phi_i)} + c(y_i, \phi_i) \right\} \quad (3.2)$$

onde θ_i é a forma canónica do parâmetro de localização, ϕ_i é um parâmetro de dispersão e $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ são funções reais conhecidas. A função $a(\cdot)$ é, por vezes, definida tal que $a(\phi_i) = \frac{\phi_i}{w_i}$, sendo w_i uma constante. $b(\cdot)$ e $c(\cdot, \cdot)$ são funções diferenciáveis que determinam o parâmetro específico da família de distribuições.

Prova-se que $E(Y) = b'(\theta_i)$ e $Var(Y) = a(\phi_i)b''(\theta_i)$. Portanto, a variância de uma distribuição, membro da família exponencial, é o produto de duas funções: uma que depende do valor médio - $b''(\theta_i)$ - e outra que depende do valor da dispersão - $a(\phi_i)$.

Os modelos lineares generalizados são considerados como uma extensão do modelo de regressão linear (clássico). De uma forma geral tais modelos são caracterizados por duas componentes:

- *Componente aleatória*: composta por uma variável resposta, Y , com n observações independentes, com valor médio μ e cuja distribuição pertence à família exponencial.
- *Componente sistemática (ou estrutural)*: composta por k variáveis explicativas que definem uma combinação linear (ou preditor linear) $\eta = \beta_0 + \sum_{j=1}^k \beta_j x_j$.

A função que relaciona as duas componentes é denominada *função de ligação* e depende da distribuição da variável resposta e do tipo de estudo que se está a fazer. São casos particulares de modelos lineares generalizados os seguintes modelos:

- Modelo de regressão linear clássico com variável resposta contínua;
- Modelo de análise de variância com distribuição normal e variável resposta contínua (ANOVA);
- Modelo de regressão logística com variável resposta binária;
- Modelo de regressão de Poisson para contagens;
- Modelo gama para variável resposta contínua.

Tabela 1. Características de alguns modelos lineares generalizados

Modelo	Componente aleatória	θ	ϕ	$a(\phi)$	$b(\theta)$	$c(y, \phi)$	Ligação canónica
Regressão linear (clássica)	Normal – $N(\mu, \sigma^2)$	μ	σ^2	σ^2	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left\{ \frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right\}$	Identidade: $\eta = \mu$
Regressão logística	Binomial – $B(n, \pi)$	$\ln\left(\frac{\pi}{1-\pi}\right)$	1	$\frac{1}{n}$	$\ln(1 + e^\theta)$	$\ln\binom{n}{ny}$	Logit: $\eta = \ln\left(\frac{\pi}{1-\pi}\right)$
Regressão de Poisson	Poisson – $P(\mu)$	$\ln(\mu)$	1	1	e^θ	$-\ln(y!)$	Logaritmica: $\eta = \ln(\mu)$
Modelo Gama	Gama – $\text{Ga}\left(\nu, \frac{\nu}{\mu}\right)$	$-\frac{1}{\mu}$	$\frac{1}{\nu}$	$\frac{1}{\nu}$	$-\ln(-\theta)$	$\nu \ln(y) - \ln \Gamma(\nu)$	Recíproca: $\eta = \frac{1}{\mu}$

Definido o tipo de MLG a ser utilizado, alguns métodos podem ser usados para estimar os parâmetros β 's da regressão, entre os quais o Método da Máxima Verosimilhança (MMV) que tem propriedades óptimas como a consistência e a eficiência assintótica. Definiremos o MMV, mais adiante, quando for aplicado à estimação dos parâmetros no modelo de regressão logística.

Em geral não é possível a obtenção de distribuições exactas para os estimadores de MV e para as estatísticas de testes usados nos MLG. Trabalha-se, portanto, com os resultados assintóticos. Fahrmer e Kaufmann (1985) referem que, para amostras grandes, o estimador de β ($\hat{\beta}$) segue assintoticamente uma distribuição normal $(k+1)$ -variada com valor médio nulo.

Quando o preditor é uma combinação de muitas variáveis, tem interesse saber qual o modelo mais parcimonioso, isto é, o modelo que envolve um mínimo de parâmetros possíveis a serem estimados e que explica bem a variabilidade da variável resposta. Existem diversos critérios para selecção de modelos e os mais utilizados são baseados na função de verosimilhança.

Seleccionado o modelo, procede-se à análise da qualidade do seu ajustamento ou seja, procede-se à avaliação da discrepância entre os valores ajustados pelo modelo e os valores observados. Existem técnicas e medidas apropriadas que permitem não só analisar a qualidade do ajustamento, como também identificar valores que não são bem explicados pelo modelo. Uma vez analisada a qualidade do modelo ajustado, procede-se, por fim, à interpretação dos resultados, em particular das estimativas dos parâmetros as quais permitem tirar conclusões sobre o estudo.

Neste estudo trabalhamos com o modelo de regressão logística pelo que vamos centrar a nossa atenção nesse modelo, sugerindo Turkman e Silva (2000) e McCullagh e Nelder (1989) para uma melhor compreensão dos modelos lineares generalizados.

2. Descrição do modelo de regressão logística

Num determinado estudo estatístico podemos estar interessados em variáveis que podem ser caracterizadas apenas pela *presença* ou *ausência* de um certo atributo - são denominadas *variáveis binárias*. As observações das variáveis binárias podem ser codificadas usando o sistema numérico (0,1). Comummente, utilizamos os termos *sucesso* (1) e *insucesso* (0) para referenciar as duas categorias. Costuma-se denominar por π a probabilidade de um indivíduo ter *sucesso* - $P(Y=1)$. Assim, uma variável binária Y pode ser definida da seguinte forma:

$$Y = \begin{cases} 0, & 1 - \pi \\ 1, & \pi. \end{cases}$$

Consideremos que k covariáveis são observadas num conjunto de n indivíduos. Cada indivíduo está associado a um vector de covariáveis (x_1, x_2, \dots, x_k) . Supomos que para o i - ésimo indivíduo dispomos do conjunto de observações $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$. Os indivíduos que partilham o mesmo vector de covariáveis formam um *padrão de covariáveis* (Turkman & Silva, 2000) ou uma *classe de covariáveis* (McCullagh & Nelder, 1989).

Considerando os dados agrupados por padrões de covariáveis, n_i o número de indivíduos pertencentes ao i - ésimo padrão de covariáveis e y_i o número de sucessos no i - ésimo padrão de covariáveis, $\frac{y_i}{n_i}$ representa a forma da resposta. O vector das dimensões dos padrões de covariáveis $\mathbf{h} = (n_1, n_2, \dots, n_k)$ é designado por *vector de índice binomial* ou *vector denominador* (McCullagh & Nelder, 1989).

O agrupamento dos dados é particularmente importante em situações em que as covariáveis são qualitativas. Quando os dados não estão agrupados por padrões de covariáveis, considera-se $n_i=1$, para todo i .

Num conjunto de n_i indivíduos pertencentes a um padrão de covariáveis, onde as observações são independentes e a probabilidade de sucesso (π_i) é constante, a variável Y_i , que conta o número de sucessos entre os n_i indivíduos, segue uma distribuição binomial de parâmetros n_i e π_i - $Y_i \cap \text{Binomial}(n_i, \pi_i)$.

Assumindo que n_i é conhecido, a função massa de probabilidade da variável Y_i tem a forma:

$$f(y_i, \pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (3.3)$$

com $y_i \in \{0, 1, \dots, n_i\}$ e $\pi_i \in [0, 1]$.

A expressão (3.3) pode ser escrita na forma:

$$\begin{aligned} f(y_i, \pi_i) &= \exp \left\{ \ln \left[\binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \right] \right\} \\ &= \exp \left\{ y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + n_i \ln(1 - \pi_i) + \ln \binom{n_i}{y_i} \right\}. \end{aligned} \quad (3.4)$$

Note que a expressão (3.4) tem a forma da função de probabilidade apresentada em (3.2), o que comprova que a distribuição binomial é membro da família exponencial. Neste caso os parâmetros da distribuição são:

- $\theta_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right)$
- $a(\phi_i) = 1$
- $b(\theta_i) = -n_i \ln(1 - \pi_i)$
- $c(y_i, \theta_i) = \ln \binom{n_i}{y_i}$

Sendo o parâmetro canónico $\theta_i = \ln \left(\frac{\pi_i}{1 - \pi_i} \right)$, então $\pi_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}}$. Consequentemente

$b(\theta_i) = n_i \ln(1 + e^{\theta_i})$ e o valor médio e a variância da distribuição binomial são tais que:

$$E(Y_i) = \mu = b'(\theta_i) = n_i \pi_i \quad (3.5)$$

$$Var(Y_i) = b''(\theta_i) = n_i \frac{e^{\theta_i}}{(1 + e^{\theta_i})^2} = n_i \pi_i (1 - \pi_i). \quad (3.6)$$

No caso particular, quando $n_i = 1$ (ou seja quando os dados não estão agrupados), $y_i \in \{0, 1\}$, estamos perante uma *distribuição de Bernoulli*. Com efeito:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1 - y_i};$$

$$E(Y_i) = \mu = b'(\theta_i) = \pi_i; \text{Var}(Y_i) = b''(\theta_i) = \pi_i(1 - \pi_i).$$

2.1. Função de ligação

Imaginemos que pretendemos relacionar uma variável binária Y_i , com probabilidade de *sucesso* π_i , com um conjunto de k covariáveis ou seja, pretendemos conhecer a relação entre a probabilidade de ocorrer *sucesso* e o vector das covariáveis $\mathbf{x} = (x_1, x_2, \dots, x_k)$. Um modelo de regressão linear correspondente seria $y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \xi$, com $E(Y_i) = \pi_i$ e $\text{Var}(Y_i) = \pi_i(1 - \pi_i)$. Assim, uma função que representaria π_i em função das covariáveis seria $E(Y_i | \mathbf{x}) = \pi_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}$. Entretanto, este resultado violaria alguns pressupostos do modelo de regressão como:

- A variância de Y_i depende de x_i ;
- Sendo que $\beta_i \in \mathcal{R}$, a igualdade $E(Y_i | \mathbf{x}) = \pi_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}$ implica que $E(Y_i | \mathbf{x})$ pode tomar qualquer valor real, o que contraria as leis de probabilidade (pois $E(Y_i | \mathbf{x}) = \pi_i \in [0,1]$). Consequentemente, exprimir π_i em função de $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$ não é apropriado.

Uma forma de ultrapassar este inconveniente é aplicar à probabilidade de *sucesso* uma função $g(\pi_i)$ linear com os parâmetros β que varia de $-\infty$ a $+\infty$ tal que:

$$g(\pi_i) = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}. \quad (3.7)$$

Aplicando uma transformação logarítmica à π_i , $g(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right)$, tem-se que

$-\infty \leq \ln\left(\frac{\pi_i}{1 - \pi_i}\right) < \infty$, sendo $g(\pi_i)$ uma função que converge para a normal. Portanto, uma

função sugerida é a **função logística** para a probabilidade π_i , $g(\pi_i)$, denotada de **logit** (π_i).

Existem ainda outras duas funções utilizadas para resolver a questão de linearidade:

- **Função Probit** – O *probit* define π como a probabilidade da variável aleatória normal padrão, Z , tomar valores inferiores a um número real ξ :

$$\pi = P(Z \leq \xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} \exp\left(-\frac{1}{2}z^2\right) dz.$$

A função pode ser representada por, $\pi = \Phi(\xi)$. Logo, uma inversa da função seria $\xi = \Phi^{-1}(\pi)$, a denominada função *probit*(π);

- **Função Log-log** – A função Log-log é dada pela expressão $\ln[-\ln(1 - \pi)]$.

A forma simples e flexível da função *logit* e a sua adaptação na interpretação de resultados com significados clínicos fazem dela a função mais utilizada em investigações ligadas às Ciências da Saúde. No nosso estudo, utilizámos o modelo de regressão logística para estudar a relação entre a nossa variável resposta e as independentes, pelo que a função *logit* é a utilizada nas próximas secções.

Considerando $\pi(x)$ o valor esperado da variável resposta dados os valores de \mathbf{x} , a função *logit* pode ser expressa da seguinte forma:

$$g(\pi(\mathbf{x})) = \ln\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) = \beta_0 + \sum_j \beta_j x_j. \quad (3.8)$$

Assim, no caso específico do modelo de regressão logística tem-se a seguinte caracterização:

- *Componente aleatória*: Dado um conjunto de n indivíduos, a variável Y segue uma distribuição binomial (logo, pertencente à família exponencial), com valor médio $E(Y | \mathbf{x}) = \mu = \pi(\mathbf{x})$ e um parâmetro de dispersão ϕ (igual a 1) que não depende dos valores de \mathbf{x} ;
- *Componente sistemática (ou estrutural)* – O valor médio de Y , $\pi(\mathbf{x})$, tem uma relação linear com o preditor $\eta = \beta_0 + \sum_{j=1}^k \beta_j x_j$ tal que $\eta = g(\pi(\mathbf{x}))$.

Resolvendo a equação (3.9) em ordem a $\pi(\mathbf{x})$, obtém-se o valor médio da variável resposta em função dos parâmetros β_j :

$$E(Y | \mathbf{x}) = \pi(\mathbf{x}) = \frac{\exp\left(\beta_0 + \sum_{j=1}^k \beta_j x_j\right)}{1 + \exp\left(\beta_0 + \sum_{j=1}^k \beta_j x_j\right)} \quad (3.9)$$

ou ainda,

$$\pi(\mathbf{x}) = \frac{\exp[g(\mathbf{x})]}{1 + \exp[g(\mathbf{x})]}. \quad (3.10)$$

No modelo de regressão linear assume-se que a variável resposta pode ser expresso como $Y = E(Y | \mathbf{x}) + \varepsilon$. No caso do modelo de regressão logística em que $Y = \pi(\mathbf{x}) + \varepsilon$, a quantidade ε pode assumir um dos dois possíveis valores - $\pi(\mathbf{x})$ ou $1 - \pi(\mathbf{x})$, tais que:

$$\begin{cases} Y = 1, P[\varepsilon = 1 - \pi(\mathbf{x})] = \pi(\mathbf{x}) \\ Y = 0, P[\varepsilon = -\pi(\mathbf{x})] = 1 - \pi(\mathbf{x}) \end{cases}$$

Assim, a variável ε segue uma distribuição de Bernoulli com valor médio $\pi(\mathbf{x})$ e variância $\pi(\mathbf{x})[1 - \pi(\mathbf{x})]$.

Tendo definido o modelo, interessa estimar os valores de β_j e atribuir-lhes as adequadas interpretações, atendendo ao fenómeno em estudo. Para tal, é importante analisar os tipos de variáveis independentes em questão. As variáveis nominais não podem ser consideradas no modelo como se fossem escalares, mas sim codificadas à custa de *variáveis dummy* (Hosmer & Lemeshow, 1989). Genericamente, se uma j -ésima covariável qualitativa tem k categorias, são necessárias $k-1$ variáveis *dummy* para a representar. Por exemplo, a covariável estado civil pode receber as categorias: Solteiro, Casado, Divorciado ou Viúvo. Admitindo-se, por exemplo, que Solteiro é o estado civil de referência, criam-se 4 variáveis *dummy*: Casado (Sim-1, Não-0), Divorciado (Sim-1, Não-1) e Viúvo (Sim-1, Não-0). As variáveis *dummy* têm de ser incluídas no modelo, e os seus coeficientes devem ser codificados de β_{jl} , com $l \in \{1, \dots, k-1\}$.

2.2. Estimação dos parâmetros – Função de verosimilhança

No modelo de regressão linear clássico, o método mais utilizado para a estimação dos parâmetros β_j é o Método dos Mínimos Quadrados (MMQ), no qual são determinados os β_j que maximizam os quadrados dos desvios de valores observados (y) dos valores ajustados (\hat{y}) baseados no modelo. Entretanto, para modelos cuja variável resposta é binária, este método não produz estimadores com propriedades estatísticas desejáveis. Assim, recorre-se ao *Método da Máxima Verosimilhança (MMV)* que estima valores para os parâmetros desconhecidos que maximizam a probabilidade de obter um conjunto de dados observados (Hosmer & Lemeshow, 1989).

A função de verosimilhança de $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$, $L(\boldsymbol{\beta} | (\mathbf{x}, y, n))$, expressa a probabilidade dos dados observados como uma função de parâmetros desconhecidos, e os seus estimadores resultantes são aqueles que mais se aproximam do resultado e é dada por:

$$L(\boldsymbol{\beta} | (\mathbf{x}, y, n)) = P(Y = y | \mathbf{x}).$$

Assumindo que as observações são independentes, para todo o modelo linear generalizado tem-se que:

$$L(\boldsymbol{\beta} | (\mathbf{x}, y, n)) = P(Y = y | \mathbf{x}) = \prod_{i=1}^n f(y_i | x_i). \quad (3.11)$$

Sendo $\pi(\mathbf{x}_i)$ o valor de $\pi(\mathbf{x})$ calculado em \mathbf{x}_i para os pares de observações (y_i, \mathbf{x}_i) com $y_i=1$, a contribuição para a função de verosimilhança é $\pi(\mathbf{x}_i)$ e, para os pares com $y_i=0$, a contribuição é $1 - \pi(\mathbf{x}_i)$. Assim, a função de verosimilhança para o modelo logístico é dada pela expressão

$$L(\boldsymbol{\beta} | (\mathbf{x}, y, n)) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} (1 - \pi(\mathbf{x}_i))^{1-y_i}. \quad (3.12)$$

Aplicando o logaritmo à expressão (3.12), obtém-se o *logaritmo da verosimilhança*, $l(\boldsymbol{\beta})$, dado pela expressão:

$$l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta} | (\mathbf{x}, y, n))] = \sum_{i=1}^n \{ y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)] \}$$

Substituindo o valor de $\pi(\mathbf{x}_i)$ dado em (3.10) e recorrendo às propriedades do logaritmo, obtemos a expressão simplificada do *logaritmo da verosimilhança* para o modelo de regressão logística:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ y_i \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right) + \ln \left(1 + \exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right) \right) \right\}. \quad (3.13)$$

Derivando a expressão (3.13) em ordem a cada elemento do vector $\boldsymbol{\beta}$, obtém-se o vector

$\mathbf{S}(\boldsymbol{\beta}) = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0}, \dots, \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} \right)$ denominado função *score*, tal que:

$$\begin{aligned} \triangleright \quad \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0} &= \sum_{i=1}^n \left[y_i - \frac{\exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right)}{1 + \exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right)} \right]; \\ \triangleright \quad \frac{\partial l(\boldsymbol{\beta})}{\partial \beta_p} &= \sum_{i=1}^n \left[y_i x_{ip} - \frac{\exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right)}{1 + \exp \left(\beta_0 + \sum_{j=1}^k \beta_j x_{ij} \right)} \right], \text{ para } p=1, \dots, k. \end{aligned}$$

Para encontrar as estimativas de máxima verosimilhança, $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \dots, \hat{\beta}_k \right)$, resolve-se o sistema que se obtém igualando cada um dos elementos do vector $\mathbf{S}(\boldsymbol{\beta})$ a zero. Uma vez que o sistema não apresenta solução analítica, para encontrar as estimativas pretendidas recorre-se a métodos iterativos (como o de *Newton-Raphson*). Prova-se, a partir das segundas derivadas de $l(\boldsymbol{\beta})$, que a função definida pelo logaritmo da verosimilhança é estritamente convexa, logo o sistema admite apenas uma solução (ou seja existe um único ponto $\boldsymbol{\beta}$ que maximiza a função $l(\boldsymbol{\beta})$). As segundas derivadas parciais são dadas pelas expressões:

$$\begin{aligned} \triangleright \quad \frac{\partial^2 l(\boldsymbol{\beta})}{\partial^2 \beta_j} &= -\sum_{i \leq n} x_{ij}^2 \pi(\mathbf{x}_i)(1 - \pi(\mathbf{x}_i)) = -\sum_{i \leq n} x_{ij}^2 \text{Var}(y_i); \\ \triangleright \quad \frac{\partial^2 l(\boldsymbol{\beta})}{\partial^2 \beta_j \beta_l} &= -\sum_{i \neq l} x_{ij} x_{il} \pi(\mathbf{x}_i)(1 - \pi(\mathbf{x}_i)) = -\sum_{i \neq l} x_{ij} x_{il} \text{Var}(y_i). \end{aligned}$$

Os termos negativos encontrados na equações acima, definem a matriz $(k+1) \times (k+1)$, $\mathbf{I}(\boldsymbol{\beta})$, denominada matriz de informação de *Fisher*. Na diagonal da inversa de $\mathbf{I}(\boldsymbol{\beta})$, encontram-se as variâncias e nos elementos fora da diagonal as covariâncias dos estimadores tal que: $\sum(\boldsymbol{\beta}) = \frac{1}{\mathbf{I}(\boldsymbol{\beta})}$. Entretanto, muitas vezes não é possível encontrar uma expressão para as componentes da inversa da matriz $\mathbf{I}(\boldsymbol{\beta})$, pelo que as variâncias e as covariâncias obtêm-se substituindo os parâmetros desconhecidos pelas suas estimativas.

Sendo $\hat{\boldsymbol{\beta}}^{(0)}$ o ponto inicial para $\boldsymbol{\beta}$, o método iterativo de Newton Raphson, permite escrever que na $(m+1)$ -ésima iteração: $\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} + (-\mathbf{I}(\hat{\boldsymbol{\beta}}^{(m)})^{-1} S^T(\hat{\boldsymbol{\beta}}^{(m)}))$.

Uma das propriedades dos estimadores de máxima verosimilhança é a propriedade de invariância - se $\hat{\theta}$ é um estimador de máxima verosimilhança para um parâmetro θ , então $g(\hat{\theta})$ é um estimador de máxima verosimilhança para $g(\theta)$. Assim, obtendo-se as estimativas dos parâmetros no modelo podem-se calcular as estimativas das probabilidades ajustadas:

$$\hat{\pi}(\mathbf{x}_i) = \frac{\exp\left(\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{ij}\right)}{1 + \exp\left(\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j x_{ij}\right)}. \quad (3.14)$$

2.3. Teste de significância dos coeficientes

Depois de se ter ajustado o modelo, pode-se testar a significância (importância estatística) das covariáveis presentes no modelo. Este processo envolve a aplicação de testes de hipóteses.

Segundo Hosmer e Lemeshow (1989), uma justificação para testar a significância do coeficiente de uma covariável em qualquer modelo está relacionada com a seguinte questão - “o modelo que inclui a covariável em questão diz mais sobre a variável resposta do que o modelo que não inclui a covariável?”. A questão pode ser formulada em termos de hipóteses:

$$H_0: \beta_j=0 \text{ versus } H_1: \beta_j \neq 0.$$

Entretanto, nalgumas situações podemos questionar se é possível reduzir um determinado modelo, excluindo um conjunto de q covariáveis sem prejudicar o ajustamento aos dados. Considerando E , o conjunto dos q índices das covariáveis que se pretende excluir do modelo, as hipóteses a testar podem ser representadas por:

$$H_0: \beta_j = 0, \forall j \in E \text{ versus } H_1: \exists j \in E: \beta_j \neq 0 .$$

Estas restrições podem ser escritas sob a forma matricial $\mathbf{C}\boldsymbol{\beta} = 0$, onde \mathbf{C} representa uma matriz $q \times (k+1)$, cujas linhas são as q linhas de uma matriz identidade $(k+1) \times (k+1)$ associadas às q covariáveis.

Para decidir se convém reduzir o modelo, compara-se os valores observados com os ajustados da variável resposta para cada um dos dois modelos - o modelo sob as restrições e o modelo corrente. Os dois modelos (o corrente e o sob H_0) que pretendemos comparar são considerados *modelos encaixados* (uma vez que um deles se obtém do outro, excluindo pelo menos um dos termos).

Existem três estatísticas para testar hipóteses referentes aos parâmetros de regressão, as quais são baseadas nas distribuições assintóticas dos estimadores de máxima verossimilhança e nas funções adequadas desses estimadores: *Estatística de razão de verossimilhanças* ou *Estatística de Wilks*, *Estatística de Wald* e *Estatística de Score (ou de Rao)*.

Estatística de razão de verossimilhanças

Esta estatística é baseada na razão entre os máximos das verossimilhanças dos dois modelos que se pretende comparar. Sendo M_C o modelo corrente e M_{H_0} o modelo sob as restrições, a estatística de razão de verossimilhanças é definida pela expressão:

$$D(M_{H_0}, M_C) = -2[l(\boldsymbol{\beta}_{M_{H_0}}) - l(\boldsymbol{\beta}_{M_C})] \quad (3.15)$$

em que os índices de β indicam o modelo em que o parâmetro é estimado.

Sob certas condições de regularidade, a estatística $D(M_{H_0}, M_C)$ segue, sob H_0 , uma distribuição assintoticamente qui-quadrado com $(k-p)$ graus de liberdade, sendo $k-p$ a diferença entre o número de parâmetros a estimar no modelo corrente e o número de parâmetros a estimar no modelo sob H_0 (Hosmer & Lemeshow, 1989). Portanto, rejeita-se H_0 , a um nível de significância α , se o valor observado de $D(M_{H_0}, M_C)$ for superior ao quantil $1-\alpha$ da distribuição χ^2_{k-p} .

Estatística de Wald

O teste de *Wald* compara o estimador de máxima verosimilhança do parâmetro, $\hat{\beta}_j$, com uma estimativa do seu desvio padrão e a estatística é definida por:

$$W = \frac{\hat{\beta}_j}{\sqrt{\text{Var}(\hat{\beta}_j)}}. \quad (3.16)$$

Sob a hipótese individual $\beta_j = 0$, a estatística W segue uma distribuição normal padrão. Assim, o valor-p bilateral correspondente ao teste de significância é a $P(|Z| > w)$, onde Z representa uma variável normal padrão.

O teste de *Wald* também pode ser realizado através de uma estatística de teste com distribuição χ^2_q quando se pretende testar uma combinação de parâmetros ($C\beta$). Desta feita, a estatística de teste é definida pela expressão:

$$\chi^2 = \left(C\hat{\beta} \right)^T \left(C\Sigma(\hat{\beta})C^T \right)^{-1} C\hat{\beta} \quad (3.17)$$

onde $\Sigma(\hat{\beta})$, representa a matriz das variâncias e das covariâncias estimadas.

Sob a hipótese nula, a estatística χ^2 tem uma distribuição assintótica qui-quadrado com q graus de liberdade. Portanto, rejeita-se a hipótese nula, a um nível de significância α , quando $\chi^2_{\text{observado}} > \chi^2_{1-\alpha, q}$.

Estatística de Score

O teste de *Score* é baseado nas derivadas do logaritmo da verosimilhança $S(\beta)$. Partindo do princípio que $S(\hat{\beta})=0$, se substituirmos $\hat{\beta}_{M_C}$ pelo novo estimador de máxima verosimilhança - $\hat{\beta}_{M_{H_0}}$ - então $S(\hat{\beta}_{M_{H_0}})$ será significativamente diferente de zero quando H_0 não é verdadeira. Assim, a estatística de *Score*, U , que analisa a diferença entre cada componente de $S(\hat{\beta}_{M_{H_0}})$ e zero, é definida pela expressão:

$$U = \frac{\frac{\partial l(\beta_{M_{H_0}})}{\partial \beta_j}}{\frac{\partial^2 l(\beta_{M_{H_0}})}{\partial^2 \beta_j}}. \quad (3.18)$$

Recorrendo a uma expressão matricial, a estatística acima pode ser definida pela expressão $U = S(\hat{\beta}_{M_{H_0}})^T [I(\hat{\beta}_{M_{H_0}})]^{-1} S(\hat{\beta}_{M_{H_0}})$. Sob H_0 verdadeira, U segue uma distribuição qui-quadrado com q graus de liberdade. Este teste é, particularmente, útil quando já se tem calculado o estimador restrito de β .

2.4. Intervalo de confiança para os parâmetros

A estimação do intervalo de confiança consiste em determinar um intervalo $[L, L_S]$ onde, com uma certa confiança (probabilidade) $1-\alpha$, esteja o parâmetro β_j desconhecido.

Um intervalo de confiança para o parâmetro β_j é baseado no seu respectivo *teste de Wald*. Sendo $z_{1-\frac{\alpha}{2}}$ o quantil da normal padrão $N(0,1)$, o intervalo de confiança para o parâmetro β_j , associado a um nível de confiança $(1-\alpha)$, é definido por:

$$\left(\hat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_j)} \right) \quad (3.19)$$

sendo $\hat{V}(\hat{\beta}_j)$ uma estimativa da variância de $\hat{\beta}_j$.

O cálculo e a interpretação do intervalo de confiança são, também, importantes para o teste de significância dos coeficientes das covariáveis. Se o intervalo de confiança para o parâmetro β_j contiver o valor zero então, com $100(1-\alpha)\%$ de confiança, pode-se afirmar que a variável independente correspondente não é importante para explicar a variável resposta.

2.5. Interpretação dos parâmetros

Após a estimação dos coeficientes de regressão (ou a selecção do “melhor” modelo), importa interpretar os seus valores. Para tal é indispensável saber o significado dos coeficientes e, consequentemente, dois conceitos importantes na regressão logística: *odds* e *odds ratio*.

Numa análise de regressão multivariada podemos encontrar a força de associação entre a variável resposta e qualquer variável explicativa, mantendo constante o efeito das outras covariáveis. Seja x_i a variável explicativa dicotómica codificada pela ausência ($x_i=0$) ou pela presença ($x_i=1$) de um atributo, consideremos $\pi_0(x)$ a probabilidade de obter *sucesso* na ausência do atributo e $\pi_1(x)$ a probabilidade de obter *sucesso* na presença do atributo (mantendo constante as outras covariáveis). A razão que representa o quociente entre a probabilidade de *sucesso* e a probabilidade de *insucesso* entre indivíduos com a presença do atributo, $\frac{\pi_1(\mathbf{x})}{1 - \pi_1(\mathbf{x})}$,

é denominada *odds* entre os indivíduos com a presença do atributo. Analogamente, $\frac{\pi_0(\mathbf{x})}{1 - \pi_0(\mathbf{x})}$ é o *odds* entre os indivíduos com a ausência do atributo.

Note que $\ln\left(\frac{\pi_0(\mathbf{x})}{1 - \pi_0(\mathbf{x})}\right) = \beta_0 + \beta_i \times 0 + \sum_{j \neq i} \beta_j x_j$ e $\ln\left(\frac{\pi_1(\mathbf{x})}{1 - \pi_1(\mathbf{x})}\right) = \beta_0 + \beta_i \times 1 + \sum_{j \neq i} \beta_j x_j$.

E assim,

$$\ln\left(\frac{\pi_1(\mathbf{x})}{1-\pi_1(\mathbf{x})}\right) - \ln\left(\frac{\pi_0(\mathbf{x})}{1-\pi_0(\mathbf{x})}\right) = \beta_l. \quad (3.20)$$

Pode-se ter um raciocínio idêntico quando a variável explicativa é numérica. Basta pensar que $\frac{\pi_a(\mathbf{x})}{1-\pi_a(\mathbf{x})}$ representa a *odds* de *sucesso* quando $x_l=a$ e $\frac{\pi_{a+r}(\mathbf{x})}{1-\pi_{a+r}(\mathbf{x})}$ representa a *odds* de *sucesso* quando $x_l=a+r$, sendo a e $a+r$ números reais com significados para a covariável em questão. Assim,

$$\begin{aligned} \ln\left(\frac{\pi_{a+r}(\mathbf{x})}{1-\pi_{a+r}(\mathbf{x})}\right) - \ln\left(\frac{\pi_a(\mathbf{x})}{1-\pi_a(\mathbf{x})}\right) &= \left(\beta_0 + \beta_l(a+r) + \sum_{j \neq l} \beta_j x_j\right) - \left(\beta_0 + \beta_l a + \sum_{j \neq l} \beta_j x_j\right) \\ &= r\beta_l \end{aligned} \quad (3.21)$$

Portanto, o coeficiente de regressão (β_l) da variável independente x_l representa a inclinação (ou a taxa de alteração) da função *logit* (ou do logaritmo do *odds*) por um acréscimo de uma unidade em x_l (considerando constantes as demais variáveis explicativas no modelo). Por conseguinte, incrementando r unidades na unidade da variável independente, a função *logit* sofre uma alteração de β_l para $r \times \beta_l$.

A razão definida pelo quociente entre os *odds* de dois grupos da variável x_l é denominada de *odds ratio* (OR_l):

$$OR_l = \frac{odds_{x_l=b}}{odds_{x_l=a}} = \frac{\frac{\pi_b(\mathbf{x})}{1-\pi_b(\mathbf{x})}}{\frac{\pi_a(\mathbf{x})}{1-\pi_a(\mathbf{x})}}. \quad (3.22)$$

Genericamente, o grupo referente ao denominador do *odds ratio* definida em (3.21) é a *categoria de referência* da variável em questão. Quando a variável x_l é dicotômica ($a=0$ e $b=1$), pode-se provar, a partir da expressão (3.20), que o *odds ratio* é igual à exponencial do parâmetro estimado:

$$OR_l = \exp(\beta_l). \quad (3.23)$$

Tal raciocínio pode ser estendido às variáveis categóricas com mais de duas categorias. Para cada categoria é calculado o *odds ratio* que toma o valor 1 para a categoria de referência, e para as restantes categorias é obtido de forma análoga descrita para as covariáveis dicotômicas (exponencial da diferença entre os logaritmos dos *odds* – categoria em questão *versus* categoria de referência). Quando a covariável x_l é numérica prova-se a partir da expressão (3.22) que o $\log(OR_l)$ é proporcional à diferença entre os valores a e b e:

$$OR_l = \exp((b - a)\beta_l). \quad (3.24)$$

O *odds-ratio* é, portanto, a razão entre o *odds* de *sucesso* entre os expostos e o *odds* de *sucesso* entre os não expostos a um determinado atributo, representando, assim, o grau de associação entre uma variável explicativa e a variável resposta, depois de controlados os efeitos de todas as demais covariáveis no modelo (motivo pelo qual é geralmente chamado de ***odds-ratio ajustado***). A expressão *odds* não possui equivalente em português, e embora alguns autores usem “*razão de chances*” em vez de *odds ratio*, optámos pela expressão original.

Um *odds ratio* igual a 1 indica que a covariável correspondente não contribui para a explicação da variável resposta. Se o *odds ratio* é igual a um valor k superior a 1, a covariável está relacionada positivamente com a variável resposta, indicando que a probabilidade de ocorrência de *sucesso* na categoria em questão é k vezes maior quando comparado com a probabilidade de ocorrência de *sucesso* na categoria de referência. Se k for um número positivo inferior a 1, a covariável está associada negativamente à variável resposta, ou seja, a probabilidade de ocorrência de *sucesso* é maior na categoria de referência. Particularmente nos estudos ligados à saúde, quando um *odds ratio* é superior a 1, a variável correspondente é considerado um factor de risco para o *sucesso* em questão.

Atendendo à propriedade de invariância dos estimadores, a partir da estimativa $\hat{\beta}_j$ pode-se obter um intervalo de confiança para o *odds-ratio*:

$$\left(\exp \left[\hat{\beta}_j - z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_j)} \right], \exp \left[\hat{\beta}_j + z_{1-\alpha/2} \sqrt{\hat{V}(\hat{\beta}_j)} \right] \right). \quad (3.25)$$

2.6. Interacção e factores de confusão

Muitos estudos estatísticos são desenhados para analisar uma associação entre a variável resposta e uma variável independente em particular, geralmente denominada *factor de exposição*. Quando o efeito do *factor de exposição* f_1 sobre a variável resposta depende dos valores assumidos por uma segunda variável explicativa f_2 estamos perante a existência de uma interacção entre as variáveis (f_1 e f_2). A variável f_2 é, habitualmente, denominado *modificador de efeito*.

Existindo interacção, o modelo de regressão deixa de ser aditivo, pois a interacção é expressa no modelo através de uma variável que representa o produto dos factores de interacção (f_1, f_2). Seja f_1 definida tal que $f_1=0$ se não exposto e $f_1=1$ se exposto, e f_2 a covariável que, por exemplo, divide a amostra segundo o género dos indivíduos, com $f_2=0$ se feminino e $f_2=1$ se masculino. Sendo $x=(f_1, f_2)$, um modelo que inclua o termo de interacção entre as duas covariáveis pode ser definida pela expressão:

$$\ln\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \lambda f_1 f_2 \quad (3.26)$$

em que λ representa o coeficiente de interacção entre o género e o *factor de exposição*. Note que os *odds-ratio* nos dois géneros são:

$$\begin{aligned} \text{➤ } OR_{\text{feminino}} &= \frac{\frac{\pi(1,0)}{1-\pi(1,0)}}{\frac{\pi(0,0)}{1-\pi(0,0)}} = \exp(\beta_1); \\ \text{➤ } OR_{\text{Masculino}} &= \frac{\frac{\pi(1,1)}{1-\pi(1,1)}}{\frac{\pi(0,1)}{1-\pi(0,1)}} = \exp(\beta_1 + \lambda). \end{aligned}$$

A hipótese de igualdade dos *odds ratio* ($OR_{\text{masculino}}=OR_{\text{feminino}}$) equivale à hipótese de inexistência de interacção ($\lambda = 0$). Portanto, a ausência de interacção entre duas variáveis independentes significa que a associação entre a variável resposta e uma das variáveis é constante nos diferentes subgrupos da outra variável. Por outro lado, a presença de interacção

implica que a associação entre a variável resposta e o *factor de exposição* muda segundo os valores do *modificador de efeito*.

A interpretação de modelos com interacção é mais complexa, especialmente quando envolvem interacções de ordem igual ou superior a dois. Se uma das variáveis da interacção for binária, uma das alternativas é fazer regressões logísticas separadas para cada valor da variável. Quando as duas variáveis são contínuas, uma das alternativas é categorizar uma delas e fazer regressões separadas (Oliveira, 2009). Embora a presença de interacções torne a interpretação do modelo mais difícil, fornece informações mais completas e dinâmicas das relações entre as variáveis.

O *efeito de confusão* é uma outra situação que podemos encontrar na relação entre variáveis explicativas. Qualquer variável f_2 (que não seja *modificador de efeito*) que esteja associada tanto à variável resposta como ao *factor de exposição* é denominada *factor de confusão*. A existência do *factor de confusão* implica que a variável f_2 não está igualmente distribuída entre os subgrupos do factor f_1 , provocando assim uma estimativa errada do efeito do *factor de exposição* sobre a variável resposta (podendo alterar tanto a força como a direcção da associação).

Seja β_1 o efeito do *factor de exposição* sobre a variável resposta estimado num modelo univariado. Para averiguar se uma covariável f_2 é *factor de confusão* basta incluí-la no modelo e verificar se o efeito de f_1 sobre a variável resposta se altera significativamente. Dois efeitos β_1 e β'_1 são considerados estatisticamente diferentes se $\frac{|\beta_1 - \beta'_1|}{\beta_1} > 0.1$.

Existem métodos próprios para eliminar o *efeito de confusão*. Um deles é através do *propensity score* (Rosenbaum & Rubin, 1983).

Sendo f um *factor de exposição*, o *propensity score* é a probabilidade de um indivíduo pertencer ao grupo dos expostos dados os valores dos factores de confusão (\mathbf{x}) - $ps(x_i) = P(f=1|x_i)$ e é estimado através do modelo de regressão logística (em que a variável resposta é o factor de exposição f e o preditor é definido pelo conjunto dos factores de confusão). O *propensity score* é, portanto, utilizado para reduzir o enviesamento decorrente das diferenças nas distribuições dos preditores e, conseqüentemente, aumentar a precisão do efeito do factor de exposição.

Existem diferentes formas de utilização do $ps(x)$ (Rosenbaum & Rubin, 1983). Uma das formas é através da estratificação, ou seja, através da divisão das observações em determinados estratos de acordo com o *propensity score*. Em cada estrato, a média do *propensity score* deve ser igual para os dois subgrupos do factor de exposição. Rosenbaum e Rubin (1983) sugerem a divisão das observações em cinco estratos alegando que geralmente tal estratificação remove 90% do enviesamento decorrente das distribuições das variáveis de confusão.

Tendo dividido as observações em estratos, o *propensity score* pode ser utilizado (como uma variável categórica) no preditor linear do modelo, em substituição dos factores de confusão, para analisar a associação entre a exposição e a variável resposta.

Para uma melhor compreensão do *propensity score* e da sua aplicação, sugerimos D'Agostinho (1998) e Rosenbaum e Rubin (1983).

2.7. Selecção e validação do modelo

Atendendo aos pressupostos do modelo de regressão múltipla, um modelo não deve conter variáveis desnecessárias e, por outro lado, todas as variáveis importantes devem constar no modelo. Um número grande de covariáveis pode fazer com que o modelo explique bem os dados mas tornando complexa a sua interpretação, e um número reduzido de covariáveis no modelo pode conduzir a uma fácil interpretação mas um ajustamento menos bom dos dados.

Antes de ajustar um modelo multivariado aos dados, Hosmer e Lemeshow (1989) sugerem uma selecção das variáveis candidatas a explicar a variabilidade da variável resposta, através de uma análise univariada (descrevendo a variável resposta em função de cada uma das variáveis independentes no estudo). Quando as covariáveis são nominais, ordinais ou contínuas com poucos valores inteiros diferentes, esta análise pode ser feita através de uma tabela de contingência da variável resposta em função dos diferentes níveis de cada covariável. Essa análise pode ser feita, também, (particularmente, quando a covariável é contínua) através de uma regressão logística simples. Hosmer e Lemeshow (1989) sugerem considerar como variáveis candidatas a explicar a variabilidade da variável resposta as que, na análise univariada, se mostraram associadas à variável resposta a um nível de significância menor ou igual a 0.25.

Uma vez definido o conjunto de covariáveis a ser incluído no modelo, é importante pensar como conseguir um modelo que inclua apenas as covariáveis importantes e as interações existentes suficientes para explicar a probabilidade de sucesso em estudo. A selecção do modelo consiste em comparar os possíveis modelos (tanto pelo número de parâmetros como pela estrutura do preditor linear) que se obtém, a partir de um modelo inicial, pela exclusão ou inclusão de parâmetros.

Pensamos ser útil, antes de prosseguir, descrever alguns modelos comumente considerados nesse processo de selecção.

Entende-se por **modelo nulo**, aquele que não inclua nenhuma covariável no preditor, ou seja a variável resposta é definida em função de β_0 . Neste caso, a variabilidade da variável resposta é explicada apenas pela componente aleatória e a média amostral não representa a estrutura dos dados. Quando o modelo tem n parâmetros (igual ao número das observações de y), diz-se **modelo saturado**. Tendo o modelo n parâmetros para cada observação, o componente sistemática explica a totalidade da variação dos dados, ajustando-os perfeitamente, pois as estimativas de máxima verosimilhança de cada π_i são as próprias observações. Existem ainda dois modelos: o **modelo minimal** e o **modelo maximal**. O primeiro contém o menor número de parâmetros necessários para o ajustamento adequado aos dados e o segundo contém o maior número de parâmetros que é possível considerar para o ajustamento. Um modelo com k parâmetros, situado entre o modelo minimal e o modelo maximal, que está a ser sujeito a investigação, é denominado **modelo corrente**.

Uma forma de seleccionar o “*melhor*” modelo é recorrendo às estatísticas de teste, descritas na secção 2.3 deste capítulo, para comparar modelos encaixados. No caso da regressão logística, o teste de razão de verosimilhanças parece ser o mais indicado sendo obtido pela diferença entre dois desvios. Entretanto, para amostras de grandes dimensões as estatísticas referidas tendem a dar resultados semelhantes, pelo que a estatística a usar poderá depender apenas da estratégia que se está a seguir. Por exemplo, quando a selecção é no sentido *forward*, em que se parte de um modelo nulo ou de um modelo minimal e se considera modelos alternativos que se obtém pela inclusão de covariáveis, a estatística sugerida é a estatística de *Score*. Quando a estratégia é no sentido contrário (sentido *backward*), em que se parte do modelo maximal e se considera alternativas de modelos que se obtém pela exclusão de covariáveis, é útil a estatística de *Wald*.

Existe, para além das estatísticas de teste citadas, um outro método utilizado para comparar ou seleccionar modelos que é o *Critério de Informação de Akaike (AIC)*. O método foi proposto por Akaike (1974) e a ideia básica é encontrar um modelo com menor valor para a função

$$AIC = -l(\beta) + 2p$$

sendo p o número de parâmetros no modelo em questão.

Sendo M o modelo corrente e S o modelo saturado, define-se como **desvio residual** do modelo corrente, $D^*(M, S)$, o valor da expressão:

$$D^*(M, S) = \frac{-2(l(\beta_M) - l(\beta_S))}{\phi}. \quad (3.27)$$

Relembrando que, no caso particular do modelo de regressão logística, $\phi=1$ e expressando o método de AIC em função do desvio residual do modelo corrente, o processo consiste em encontrar um modelo que minimiza a quantidade $AIC = D^*(M, S) + 2p - 2l(\beta_S)$.

Existem alguns métodos automáticos que podem ser utilizados na decisão de incluir e excluir variáveis do modelo. O método *stepwise* é dos mais utilizados. Essencialmente, este processo calcula uma sequência de equações, adicionando (sentido *forward*) ou excluindo (sentido *backward*), a cada passo, uma variável independente.

O critério para adicionar ou excluir uma variável pode ser previamente estabelecido (como por exemplo, os valores $-p$ associados e o AIC). Por exemplo, usando o critério de informação de Akaike (AIC), a cada passo do *stepwise* é verificado se a inclusão ou a exclusão de uma variável independente faz com que o AIC diminua.

Este processo inicia-se com o ajustamento do modelo nulo e adiciona uma variável de cada vez, seleccionando em primeiro lugar a que é mais importante (de acordo com o critério estabelecido). É efectuada uma comparação entre o modelo sem e o modelo com a covariável incluída. Caso não seja importante a inclusão da covariável no modelo inicial, o processo termina aqui. Caso contrário, partindo deste novo modelo (que inclui a covariável seleccionada), introduz-se individualmente as demais covariáveis, e cada um destes novos modelos é comparado com o modelo anterior e com base no critério estabelecido selecciona-se

a próxima covariável a ser incluída. Sempre que o preditor do modelo corrente tenha duas ou mais covariáveis analisa-se a exclusão individual de cada uma delas (ou seja, caso alguma variável tenha perdido significância, com a inclusão de uma nova variável, ela é excluída do modelo). O processo continua até que não haja mais covariáveis importantes para entrar no modelo.

Quando o grau de importância é baseado no nível de significância do teste de razão de verossimilhanças entre modelos que incluem e excluem as covariáveis, como a variável mais importante pode não ser significativa do ponto de vista estatístico, impõe-se um limite superior P_E (pertencente ao intervalo $[0.15, 0.25]$) a fim de atrair candidatos importantes à entrada. Para a remoção das covariáveis, os níveis de significância são comparados com um limite inferior P_S , e as covariáveis para as quais o nível de significância é superior a P_S são candidatas à remoção.

Após selecção do modelo, pode-se excluir do modelo as variáveis que, apesar de seleccionadas pelo método, não mostraram ser estatisticamente significativas, atendendo ao nível de significância estabelecido, e a inclusão de interacções pode ser testada, individualmente, de entre aquelas de interesse no estudo.

De referir que a selecção de um modelo de regressão logística deve ter em conta não só a selecção estatística de modelos como também o bom senso. Particularmente, nas investigações ligadas à saúde, covariáveis que não mostram ser estatisticamente significativas podem ser consideradas biologicamente importantes. Portanto, experiências ou algumas considerações teóricas sobre o fenómeno em estudo podem ajudar o investigador a seleccionar o conjunto de covariáveis para descrever o fenómeno.

2.8. Qualidade de ajustamento do modelo

Após uma escolha cuidadosa de um modelo, é possível que o seu ajustamento aos dados não revele resultados satisfatórios por revelar uma certa discrepância entre os valores observados e os valores ajustados, ou ainda por revelar um ou mais valores observados discordantes em relação aos demais. A ocorrência de uma ou de outra destas duas falhas pode ter causas diferentes, embora haja uma combinação das diferentes causas, como: aplicação de uma escala

errada a, pelo menos, uma das variáveis, uma leitura (ou transcrição) mal feita dos dados, ou ainda um factor não controlado que provoca a obtenção de valores discordantes.

Os métodos para a análise da qualidade do modelo podem ser formais - baseados nas medidas de discrepância - ou informais – baseados numa análise gráfica dos resíduos.

Medidas de qualidade de ajustamento do modelo

Logicamente que não é de esperar que os valores ajustados pelo modelo e os observados sejam exactamente iguais, logo uma discrepância pequena pode ser tolerável, mas uma discrepância grande já implica que o modelo ajustado precisa ser revisto. Uma forma de medir a discrepância entre os valores observados e os ajustados por um modelo linear generalizado é através do valor do **desvio residual** do modelo corrente, definido pela expressão (3.27). Para o caso específico da regressão logística, utilizando as equações (3.10) e (3.13) tem-se:

$$D^*(M, S) = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right) \right]. \quad (3.28)$$

No caso de dados binários onde $n_i=1$, os únicos valores possíveis para y_i são 1 e 0, e, consequentemente, para o modelo saturado, as parcelas $y_i \ln(y_i)$ e $(1 - y_i) \ln(1 - y_i)$ são iguais a zero (ou seja $l(\beta_s) = 0$). Portanto, $D^*(M, S)$ depende apenas das probabilidades de sucesso ajustadas, podendo ser definido pela seguinte expressão:

$$D = D^*(M, S) = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}(x_i)}{1 - \hat{\pi}(x_i)} \right) + (1 - y_i) \ln \left(1 - \hat{\pi}(x_i) \right) \right]. \quad (3.29)$$

Sendo D definido apenas em função das probabilidades ajustadas, não é informativo quanto à discrepância entre valores observados (ajustados pelo modelo saturado) e os ajustados pelo modelo corrente.

Nelder e Wedderburn (1972) sugerem o desvio residual como uma medida de discrepância. Sob a hipótese nula do modelo corrente ser correcto, $D \sim \chi_{n-p}^2$, sendo p a dimensão do vector de

parâmetros no modelo. Ou seja, se D for inferior ao valor crítico $\chi^2_{n-p, 1-\alpha}$, pode-se concluir que o modelo corrente se ajusta bem aos dados.

Note-se que o desvio é sempre maior que zero e decresce à medida que as covariáveis vão entrando do modelo, sendo nulo quando o modelo é saturado. Procuram-se, portanto, modelos simples com um desvio moderado, capaz de explicar os dados.

Outra medida para analisar a qualidade de ajustamento do modelo é a **estatística de Pearson generalizada**. Seja L o número possível de combinações das k covariáveis no estudo, m_l o número de indivíduos da l -ésima classe de covariáveis e n_{li} o número de indivíduos com sucesso na l -ésima classe de covariáveis. Note-se que $\sum_{l=1}^L n_{li} = n_1$ representa o número de indivíduos com sucesso na amostra em estudo. Sendo $\hat{\pi}_l$, a probabilidade de sucesso ajustada na l -ésima classe de covariáveis, define-se **resíduo de Pearson** como sendo a diferença entre o valor observado e valor ajustado dado pela expressão:

$$R_l^P = \frac{n_{li} - m_l \hat{\pi}_l}{\sqrt{m_l \hat{\pi}_l (1 - \hat{\pi}_l)}}. \quad (3.29)$$

Note-se que, quando os dados não estão agrupados ($l=i$) $R_i^P = \frac{y_i - \hat{\pi}(x_i)}{\sqrt{\hat{\pi}_i (1 - \hat{\pi}_i)}}$. A estatística de *Pearson generalizada* é dada pela soma dos quadrados dos resíduos de Pearson:

$$\chi^2_P = \sum_{l=1}^L (R_l^P)^2. \quad (3.30)$$

Prove-se que χ^2_P aproxima-se, assintoticamente, de um qui-quadrado como $L-p$ graus de liberdade. Embora a estatística de Pearson tenha uma interpretação mais directa, a propriedade aditiva da *função desvio*, faz com que a estatística D seja preferível. Entretanto, a estatística D não se mostra apropriada quando $n \rightarrow \infty$ e $m_l \pi_l (1 - \pi_l)$ toma valores muito pequenos.

Hosmer e Lemeshow (1989) sugerem uma estatística alternativa para avaliação da qualidade de ajustamento. Os referidos autores sugerem a formação de g grupos (preferencialmente 10 grupos com a mesma dimensão) e comparam o número observado com o número esperado de

sucessos em cada grupo. O primeiro grupo g_1 deve conter os n_1 elementos correspondentes às n_1 primeiras menores probabilidades ajustadas ($\pi_1 \leq \pi_2 \leq \dots \leq \pi_{n_1}$). O segundo grupo g_2 deve conter as próximas n_2 menores probabilidades ajustadas $\pi_{n_1+1} \leq \pi_{n_1+2} \leq \dots \leq \pi_{n_1+n_2}$. E assim sucessivamente, o último grupo deverá conter as n_g maiores probabilidades ajustadas.

A estatística de *teste de Hosmer - Lemeshow* é definida por:

$$G_{HL}^2 = \sum_{j=1}^g \frac{(o_j - e_j)^2}{e_j \left(1 - \frac{e_j}{n_j}\right)} \quad (3.31)$$

em que o_j é o número observado de sucessos no j -ésimo grupo, n_j é o número de indivíduos no j -ésimo grupo e $e_j = \sum_{i=1}^{n_j} \hat{\pi}_{ji}$ é o número esperado de sucessos no j -ésimo grupo (sendo $\hat{\pi}_{ji}$ a probabilidade estimada correspondente à j -ésima observação do grupo j).

Existindo empates, os autores sugerem que os indivíduos com as mesmas configurações de covariáveis sejam colocados em grupos adjacentes, a fim de garantir que estes não tenham muita diferença em termos de dimensões. Hosmer e Lemeshow verificaram que, sob a condição do modelo se ajustar bem aos dados, a estatística G_{HL}^2 aproxima-se de um qui-quadrado com $(g-2)$ graus de liberdade. Assim, para valores G_{HL}^2 inferiores ao valor crítico $\chi_{g-2, 1-\alpha}^2$ conclui-se que o modelo se ajusta bem aos dados.

É importante notar que o facto de um modelo ter bom ajustamento e uma boa capacidade explicativa não implica que o seu desempenho na classificação dos indivíduos (com *sucesso* ou com *insucesso*) seja bom. Assim, um outro processo muito utilizado para avaliar o modelo é a construção da curva ROC (*Receiver Operating Characteristic*).

Para avaliar o desempenho do modelo pode-se começar por definir uma regra de decisão baseada nas probabilidades ajustadas pelo modelo. Um exemplo para esse limiar é 0.5: se a probabilidade ajustada de um indivíduo ter *sucesso* for maior que 0.5, o indivíduo é classificado como tendo *sucesso*, caso contrário o indivíduo é classificado como tendo *insucesso*. Tendo a

regra definida, pode-se dispor as observações numa tabela de contingência 2x2 (Tabela 2) que mostra como os indivíduos foram classificados:

Tabela 2. Tabela de contingência para testes de diagnóstico

Classificados	Observados	
	Sucesso	Insucesso
Sucesso	a- (Verdadeiro positivo)	b- (Falso positivo)
Insucesso	c- (Falso negativo)	d- (Verdadeiro negativo)

A partir desta tabela podemos obter as medidas importantes de desempenho do modelo: a *sensibilidade* e a *especificidade*. A *sensibilidade* (ou taxa dos verdadeiros positivos) é a fracção dos indivíduos classificados com sucesso entre todos os indivíduos que têm sucesso, enquanto a *especificidade* (ou taxa dos verdadeiros negativos) é a fracção dos indivíduos classificados com *insucesso* entre todos os indivíduos que têm *insucesso*.

São calculadas a *sensibilidade* e a *especificidade* em toda a amostra, considerando diferentes pontos de corte (limiares) do modelo. A curva ROC é baseada nas duas proporções: a fracção dos verdadeiros positivos entre os indivíduos com sucesso - $a/(a+c)$ - e a fracção dos falsos positivos entre os indivíduos com *insucesso* - $b/(b+d)$.

Num modelo perfeito (que classificasse correctamente todos os indivíduos), a curva ROC passaria pelo canto superior esquerdo de um plano unitário em que nas ordenadas se representam os valores da *sensibilidade* e nas abcissas os valores de 1- *especificidade*, pois a *sensibilidade* e a *especificidade* seriam ambas iguais a 100%. Portanto, quanto mais esta curva se afastar deste canto pior é o desempenho do modelo na classificação dos indivíduos. A área sob a curva ROC mede, portanto, a capacidade de discriminação do modelo. Hosmer e Lemeshow (1989) sugerem uma regra geral para a avaliação do modelo segundo a área sob a curva ROC:

- Se a área variar entre 0.7 e 0.8 o modelo tem um poder discriminante aceitável;
- Se a área variar entre 0.8 e 0.9 o modelo tem um bom poder discriminante;
- Se a área for superior a 0.9 o modelo tem um poder discriminante excelente.

A área sob a curva ROC pode ser traduzida como a proporção de indivíduos correctamente classificados pelo modelo independentemente do valor escolhido para o limiar de decisão.

Diagnóstico do modelo

Uma das desvantagens da utilização das estatísticas para a análise da qualidade de ajustamento do modelo é que nenhuma delas dá qualquer indicação da direcção a tomar a partir do modelo ajustado quando é rejeitada a hipótese do modelo se ajustar bem aos dados. Por isso, a análise dos resíduos, através de uma representação gráfica, é essencial para analisar a qualidade do ajustamento do modelo e identificar os pontos influentes ou mal ajustados. Com as medidas de influência podemos verificar até que ponto, excluir ou incluir uma observação fora do padrão comum, terá impacto nos parâmetros de regressão e nas estatísticas de qualidade do ajustamento. Para uma análise adequada, convém que os resíduos tenham a mesma variância. Assim há autores sugerem a utilização de resíduos padronizados (Williams, 1987).

No estudo da regressão linear clássica é definida a matriz \mathbf{H} (*hat matrix*) cujos elementos h_{ij} representam a influência de y_j sobre \hat{y}_i . Sendo o modelo definido por $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\xi}$, no caso da regressão logística, \mathbf{H} é dada pela expressão

$$\mathbf{H}=\mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2}$$

onde \mathbf{W} representa uma matriz diagonal de ordem n cujo i -ésimo elemento é definido por $w_{ii} = \pi_i(1 - \pi_i)$. Assim, os elementos principais de \mathbf{H} , denominados *leverage*, são dados pela expressão $h_{ii} = w_{ii}x_i^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}x_i^T$, onde x_i representa a i -ésima linha da matriz das covariáveis. Prova-se que $\sum_{i=1}^n h_{ii} = p$, sendo p o número de parâmetros no modelo e que

$0 \leq h_{ii} \leq 1$ e valores elevados de h_{ii} ($h_{ii} > \frac{2p}{n}$) correspondem a pontos extremos (observações atípicas). Mas, h_{ii} não depende apenas de x_i como também dos estimadores dos parâmetros do modelo através de w_i . Assim, pontos extremos não correspondem necessariamente, apenas, a valores elevados do *leverage*.

Assumindo que $E(r_i) = 0$ (r_i dado por 3.29), um resíduo padronizado é definido por:

$$R_l^* = \frac{y - \hat{\pi}_l}{\sqrt{\text{Var}(Y_l)(1 - h_l)}}.$$

Sendo $\text{Var}(Y_l | \mathbf{x}_l) = m_l E(Y_l | \mathbf{x}_l)[1 - E(Y_l | \mathbf{x}_l)] = m_l \pi_l (1 - \pi_l)$, a variância de um resíduo é aproximadamente igual a $m_l \pi_l (1 - \pi_l)(1 - h_l)$. Partindo das definições do resíduo de Pearson (em 3.30) e do desvio residual (em 3.29), pode-se definir o resíduo de Pearson (R^{*P}) e o desvio residual padronizados (R^{*D}):

➤ **Resíduo de Pearson padronizado** - $R^{*P}_l = \frac{n_{1l} - m_l \hat{\pi}_l}{\sqrt{m_l \pi_l (1 - \pi_l)(1 - h_l)}};$

➤ **Desvio residual padronizado** – O desvio residual pode ser, também, definida a partir dos componentes d_i – contribuição da i -ésima observação para a função desvio. Um valor grande de d_i , mostra que a i -ésima observação é mal ajustada pelo modelo. As raízes quadradas das componentes do desvio exibem propriedades de uma função que padroniza a distribuição dos resíduos igualando as variâncias. Assim, o desvio residual padronizado correspondente à i -ésima observação é definido pela expressão:

$$R^{*D}_i = \pm \sqrt{\frac{d_i}{1 - h_{ii}}} = \pm \sqrt{\frac{2}{1 - h_{ii}}} \left\{ y_i \ln \left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \right) + (1 - y_i) \ln(1 - \pi_i) \right\}^{1/2}.$$

O sinal \pm depende da diferença $y_i - \hat{\pi}_i$. Williams (1987) sugere a utilização destes resíduos padronizados no diagnóstico de modelos lineares generalizados, com preferência do desvio residual, uma vez que este tem apresentado propriedades similares às do resíduo padronizado do modelo de regressão linear.

Como uma observação influente não tem necessariamente um resíduo elevado, e a sua modificação ou exclusão do modelo produz alterações significativas nas estimativas dos parâmetros do modelo, existem outras estatísticas de diagnóstico para identificar informações influentes. Uma delas é a designada DFBETA que é dada pela expressão:

$$\Delta \beta_{(j)} = \frac{\hat{\beta}_j - \hat{\beta}_{(i)j}}{\sqrt{\text{var}(\hat{\beta}_j)}} \quad (3.32)$$

em que $\hat{\beta}_j$ representa a estimativa do j -ésimo elemento do vector β obtida na amostra com o indivíduo i e $\hat{\beta}_{(i)j}$ é a estimativa do j -ésimo elemento do vector β obtida na amostra sem o indivíduo i . O valor de $\Delta\beta_{(j)}$ reflete, portanto, a influência da observação (y_i, \mathbf{x}_i) no valor estimado do coeficiente de x_j . Um valor de DFBETA relativamente elevado em valor absoluto indica que a observação influencia a estimativa de $\hat{\beta}_j$. Geralmente, se $|\Delta\beta_j| > 1$ para amostras pequenas ou se $|\Delta\beta_j| > 2/(n)^{1/2}$ para amostras grandes, a correspondente observação é considerada influente. No entanto, quando temos um grande numero de variáveis explicativas a análise dos valores de DFBETA torna-se cansativo.

Para medirmos a influência geral da observação i na estimativa dos coeficientes de regressão utiliza-se a *distância de Cook* aproximada (Hosmer & Lemeshow, 1989) definida, a partir da componente do desvio residual, por:

$$D_i^C = \frac{d_i}{(1 - h_{ii})}. \quad (3.33)$$

Ou ainda a partir do resíduo de Perason, $\Delta\chi_i^2 = \frac{R_i^P}{(1 - h_{ii})}$.

Devem ser analisadas as observações cuja D_i^C seja superior a uma unidade.

Existem várias representações gráficas propostas para o diagnóstico de modelos lineares generalizados. Passamos a referir as mais usadas no caso específico da regressão logística:

1. $(R_i^{*P})^2$ versus $\hat{\pi}_i$ e $(R_i^{*P})^2$ versus h_{ii}
2. $\Delta\beta_{(j)}$ versus $\hat{\pi}_i$ e $\Delta\beta_{(j)}$ versus h_{ii}
3. D_i^C versus $\hat{\pi}_i$ e D_i^C versus h_{ii}

Tais representações são particularmente úteis para identificar casos indevidamente influentes.

A influência geral de cada caso na estimação dos coeficientes pode ser analisada num gráfico de bolhas em que os símbolos são círculos proporcionais ao valor da influência.

CAPÍTULO 4: METODOLOGIA

Neste capítulo, procederemos à exposição das metodologias utilizadas na nossa abordagem, recorrendo à caracterização do tipo de estudo, à definição da população, às técnicas de amostragem, à caracterização das variáveis, aos instrumentos e procedimentos estatísticos utilizados.

1. Tipo de estudo

Cabo Verde é um país formado por dez ilhas, sendo apenas nove habitadas. Hoje, o ensino secundário é ministrado em todas as nove ilhas, tem duração de seis anos (do 7º ao 12º ano de escolaridade) e está organizado em três ciclos de dois anos cada - o 1º ciclo, correspondendo aos 7º e 8º anos de escolaridade, os 2º e 3º ciclos, correspondendo, respectivamente, aos 9º e 10º anos e aos 11º e 12º anos de escolaridade. Dada a disparidade geográfica do país (e também ao limitado tempo disponível para o estudo), limitámos a nossa abordagem apenas a uma das ilhas – São Vicente. Portanto, trata-se de um *estudo de caso* (estudando apenas um subconjunto do conjunto de todos os alunos do ensino secundário do país).

2. População do estudo e técnica de amostragem

Em São Vicente existem cinco escolas secundárias públicas, a saber: *Escola Industrial e Comercial do Mindelo*, *Escola Salesiana de Artes e Ofícios de São Vicente*, *Escola Secundária Jorge Barbosa*, *Escola Secundária José Augusto Pinto* e *Liceu Ludgero Lima*. A população alvo do estudo é constituída por todos os alunos do ensino secundário (público) da ilha, perfazendo um total de 6867 alunos (segundo os dados conseguidos junto das escolas). As duas maiores proporções desta população estudantil estão nas escolas secundárias Jorge Barbosa e José Augusto Pinto, cabendo à primeira 28.9% e à segunda 28.2%. As escolas Liceu Ludgero Lima, Escola Salesiana de Artes e Ofícios de São Vicente e Escola Industrial e

Comercial do Mindelo participam com 20.6%, 12.0% e 10.3%, respectivamente, da população considerada.

Inicialmente, pensámos estudar uma amostra de 700 alunos que representaria, aproximadamente, 10% da população. Entretanto, pensando nos custos e no tempo para a organização dos dados, reduzimos a dimensão da amostra para 500 (7.5% da população). Dadas as diferenças (em termos populacionais) entre as escolas, e pretendendo ter uma representatividade de todas as escolas na amostra, optámos por uma amostra estratificada (com participação proporcional), tal que a dimensão de cada estrato (escola) na amostra é proporcional à dimensão do estrato na população. Igualmente, a dimensão de cada ciclo na sub-amostra escola é proporcional à dimensão de cada ciclo na população escola. A técnica aplicada permitiu-nos obter a amostra estratificada apresentada na tabela seguinte (em que a percentagem da escola, dada em cada linha, corresponde à porção do respectivo n no valor Total da linha).

Tabela 3 . Representação dos diferentes ciclos e escolas secundárias na amostra

<i>Nível de ensino</i>	<i>E.I.C.Mindelo</i>		<i>J. Barbosa</i>		<i>José A. Pinto</i>		<i>L. L. Lima</i>		<i>Salesianos</i>		<i>Total</i>
	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>N</i>	<i>%</i>	
1º Ciclo	2	0.9	70	30.6	75	32.7	47	20.5	35	15.3	229
2º Ciclo	16	10.3	41	26.3	38	24.3	34	21.8	27	17.3	156
3º Ciclo	30	26.1	37	32.2	23	20.0	25	21.7	0	0%	115
Total	48	9.6	148	29.6	136	27.2	106	21.2	62	12.4	500

3. Instrumentos e procedimentos estatísticos

Para recolher os dados, desenvolvemos um questionário (ver anexo B), tendo em conta alguns já aplicados (como o de Breda, 1996 e o AUDIT da OMS por Babor *et al.*, 1992), procurando uma aproximação à realidade cabo-verdiana, e tendo uma apreciação crítica de um pedagogo e psicólogo. Este questionário, com 24 questões (na maioria, questões múltiplas), contempla 4 partes: **I. Perfil do aluno** – dados referentes à idade, ao sexo e ao ano de escolaridade do inquirido; **II. Informação familiar** – dados referentes ao agregado familiar, grau de escolaridade e profissão dos pais do inquirido; **III. Percorso escolar** – dados referentes ao número de reprovações por ano de escolaridade e percepção do aluno relativamente a alguns

atributos pessoais condicionantes ao (*in*)sucesso escolar; **IV. Comportamentos e atitudes face ao consumo de álcool** – dados referentes ao primeiro consumo, à frequência de consumo, à quantidade habitualmente consumida, aos dias de semana para o consumo, à frequência do consumo *binge* nos últimos 30 dias antes da aplicação do questionário, à frequência de embriaguez, à bebida mais consumida, às motivações de consumo, aos aspectos prejudicados pelo consumo, ao consumo entre os amigos mais próximos, ao consumo regular e problemático na família e à frequência das consequências sofridas pelo consumo de bebidas alcoólicas.

Inicialmente, o questionário foi aplicado a um grupo de 10 alunos, numa das escolas participantes, com o objectivo de ajustar os conteúdos dos itens à compreensão dos respondentes e considerar aspectos para gestão do tempo das instruções. Como resultado desta aplicação foi necessário fazer uma ligeira alteração relativamente a uma das questões.

A aplicação do questionário realizou-se, com a autorização dos directores das escolas participantes, durante a semana de 19 a 23 de Março do corrente ano (um dia em cada escola), durante os primeiros 30 minutos de administração de aula. Em cada período de leccionação, (período de manhã e período de tarde) e em cada nível de ensino, as turmas a participar foram seleccionadas aleatoriamente. Na sala de aula, e antes de aplicar o questionário, foi explicado em que âmbito o questionário estava a ser aplicado, a importância do estudo e o objectivo do mesmo. Sempre que o número de questionários foi inferior ao número de alunos da turma, os alunos foram seleccionados aleatoriamente, deixando sempre claro a possibilidade do aluno recusar, caso não quisesse, participar como respondente (nenhum aluno se negou a responder).

O questionário foi aplicado a 500 alunos, nas cinco escolas (públicas) da ilha de São Vicente, que leccionam o ensino secundário. Devido à incoerência observada nalgumas respostas e a algumas não - respostas importantes para o estudo, retirámos 10 questionários, ficando a amostra com uma dimensão igual a 490 (7.3% da população).

Para a análise dos dados recorreremos a dois *softwares* estatísticos (o *SPSS* e o *R*). A análise descritiva do percurso escolar e do padrão de consumo da amostra foi feita recorrendo ao programa *SPSS - Statistical Package for the Social Sciences – versão 19.0*. Para analisar a associação entre variáveis recorreu-se ao teste do Qui-quadrado e foi usado o nível de significância 0.05. Para testar a hipótese da variável idade seguir uma distribuição normal recorreu-se ao teste de ajustamento de *Kolmogorof – Smirnof*. Em alternativa ao teste *t* de

Student, utilizou-se o teste não paramétrico de *Mann-Whitney*, para testar a hipótese de igual distribuição da variável idade nos dois géneros.

Para realizar o estudo analítico – estudar a relação entre o consumo de álcool e o insucesso escolar - recorreremos ao *software R – versão 2.10.0*. O consumo de álcool (factor principal em estudo) foi analisado, separadamente, como uma variável dicotómica (Consumidor/Não Consumidor) e como uma variável com 4 categorias (Não consumidor, Consumidor ocasional, Consumidor menos frequente e Consumidor frequente). No total, foram construídos três modelos multivariados, sendo que o primeiro foi construído para descrever o insucesso escolar em função das covariáveis *tradicionais*, o segundo para descrever o insucesso escolar em função da exposição ao consumo de álcool e um terceiro modelo para analisar o insucesso escolar em função da frequência de consumo (ao lado das variáveis *tradicionais*).

No modelo de regressão logística multivariada foram incluídas apenas as variáveis independentes que se mostraram associadas ao insucesso escolar com um nível de significância menor ou igual a 0.25 (como sugerem Hosmer e Lemeshow, 1989). Previamente à realização da análise multivariada, procedeu-se a uma análise univariada das variáveis a incluir no modelo, através do ajustamento de modelos de regressão simples.

Para a selecção de cada modelo, recorreu-se ao método de menor AIC sugerido por Akaike (1974) através do processo *stepwise*. O teste de significância dos coeficientes de cada modelo foi efectuado com recurso ao Teste de *Wald*, com um nível de significância igual a 0.05. Após a selecção das variáveis explicativas foi testada a existência de interacção entre variáveis e a existência de factores de confusão. Para reduzir o viés relacionado com as variáveis de confundimento recorreremos ao método da *propensity score*. A qualidade de ajustamento de cada modelo seleccionado foi analisada através do teste de Hosmer & Lemeshow. A área sob a curva ROC foi utilizada para medir a capacidade de cada modelo para discriminar os alunos com insucesso escolar *versus* os sem insucesso escolar.

4. Variáveis consideradas

Neste estudo, a variável dependente é o *insucesso escolar*. Em Cabo Verde, para cada ano de escolaridade existe um conjunto mínimo de objectivos que o aluno deve atingir para que este possa passar para o ano de escolaridade seguinte. A “impossibilidade” do aluno em atingir tais

objectivos (expressa em três ou mais disciplinas com deficiências ou negativas) implica a retenção no ano e define (pela instituição escolar) o insucesso escolar. Entretanto, o aluno que não tenha tido reprovação no ano de escolaridade, poderá ter insucesso nalgumas disciplinas (no máximo duas), o que revela uma certa limitação para afirmar que este aluno tenha tido sucesso escolar no ano escolar em questão. Assim, evitando recorrer ao termo “sucesso escolar”, neste estudo a variável dependente *insucesso escolar*, é uma variável dicotómica que recebe as categorias “Apresenta insucesso” ou “Não apresenta insucesso”. A primeira categoria define-se para o aluno que apresentar no mínimo uma reprovação no contexto considerado e a segunda categoria para o caso contrário.

O objectivo geral deste estudo é, como já referimos, analisar o efeito do consumo de álcool no insucesso escolar dos alunos do ensino secundário. Assim, interessa-nos considerar o insucesso escolar como uma ocorrência após o início do consumo de álcool. Por exemplo, um aluno que apresenta insucesso escolar apenas no 9º ano e inicia o consumo de álcool no 10º ano de escolaridade é considerado um “consumidor” que “apresenta retenções”, e isto representa uma situação em que o insucesso escolar ocorre antes do início de consumo. Assim, e como o tempo médio de consumo observado na amostra é de 2.66 anos, no sentido de não considerar (ou de pelo menos minimizar o número de) situações do tipo referido, a variável dependente a considerar no estudo analítico é o insucesso escolar nos dois últimos anos de escolaridade frequentado pelo aluno (e não o insucesso em todo o percurso escolar secundário).

Sendo várias as variáveis independentes, optámos por apresentá-las agrupadas da seguinte forma:

- Variáveis sócio- demográficas: neste grupo temos as variáveis: idade, género, tipo de agregado familiar, estatuto profissional do pai e estatuto profissional da mãe.

O tipo de agregado familiar é uma variável que recebe as categorias: Nuclear – família com presença de ambos os progenitores; Monoparental – família com presença de apenas um dos progenitores; Reconstituída – família com presença de um dos progenitores e do seu cônjuge (padrasto ou madrasta do inquirido); Outros – família com ausência de ambos os progenitores. O reduzido número de estudos sobre a classe social em Cabo Verde dificultou a definição da variável baseada na profissão dos pais. A variável estatuto profissional (tanto do pai como da mãe) é categorizada, baseada no estudo (embora não recente) de Afonso (2002), da seguinte forma: Classe

Independente/Não Remunerada; Classe Dirigentes/Técnicos; Classe Executante/Administração Pública; Classe de Profissões Elementares; Reformados; Desempregados. (As duas últimas categorias não constam do estudo da referida autora).

- Variáveis escolares: conjunto de variáveis que inclui: o ano de escolaridade do inquirido, o percurso escolar no ensino básico (com retenções/sem retenções), o nível de escolaridade dos pais (analfabeto, ensino básico, ensino secundário ou formação superior) e os atributos condicionantes do sucesso escolar - gostar de estudar, faltar muito às aulas, bom ambiente familiar, interesse familiar para os estudos, acompanhamento familiar nas tarefas escolares, dificuldades em acompanhar as matérias, considerar elevadas as exigências dos professores e considerar importante os estudos para o futuro.

Relativamente às variáveis consideradas atributos condicionantes do insucesso escolar, estas recebem os valores “Sim” ou “Não”, conforme o respondente se considera com ou sem o atributo em questão.

- Primeiro consumo de álcool: refere-se à idade, ao local e à companhia na primeira experiência com o álcool;
- Consumo frequente de álcool: neste grupo estão as variáveis referentes ao consumo (Consumidor/Não consumidor), ao local de consumo, à companhia habitual de consumo, aos dias da semana para o consumo, à frequência, à quantidade habitualmente consumida e às motivações de consumo. Incluem-se também a frequência de consumo binge nos últimos 30 dias que antecede a aplicação do questionário, a frequência de embriaguez na vida, a bebida mais consumida, os aspectos prejudicados pelo consumo, o consumo entre os amigos mais próximos, o consumo regular e problemático na família e a frequência da ocorrência das consequências após o consumo de álcool.

5. Limitações do estudo

Como qualquer estudo, onde a colheita de dados se faz através de questionários, o nosso estudo apresenta algumas limitações. Uma delas refere-se à questão que se possa levantar relativamente à fidelidade dos dados, uma vez que são colhidos entre adolescentes, que podem negar o consumo quando realmente consomem ou então inflacionar os valores de consumo relatando, orgulhosamente, que consomem quando realmente não o fazem (ou pelo menos não o

fazem no padrão assumido). Entretanto, se por um lado seria possível conseguir através de dados oficiais algumas observações (como o agregado familiar, as faltas às aulas e as retenções, entre outras), por outro lado, nenhuma outra pessoa ou instituição estaria em condições de nos oferecer uma informação mais real do consumo de álcool do que o próprio consumidor.

Alguns estudos têm demonstrado que os relatos sobre o consumo de álcool ao longo do tempo são fiáveis desde que a confidencialidade, o anonimato e a privacidade dos respondentes sejam requisitos da colheita de dados (Breda, 1996) e que a aplicação (do questionário) não ultrapasse os 45 minutos (Ferreira & Campos, 2001). Neste estudo procuramos respeitar tais requisitos e o tempo de aplicação não atingiu os 45 minutos.

Outra limitação prende-se com o facto de não ser possível considerar os alunos que abandonaram o ensino secundário devido ao número de retenções obtidas no secundário. Se considerássemos tais alunos ou se considerássemos como *sucesso* a ausência de negativas nas classificações finais, provavelmente os resultados seriam diferentes.

CAPÍTULO 5: APRESENTAÇÃO DOS RESULTADOS

Neste capítulo apresentamos os dados obtidos e os tratamentos realizados após a sua recolha. Num primeiro momento caracterizamos o percurso escolar e o padrão de consumo de álcool entre os inquiridos em função de algumas variáveis (género e faixa etária). Com o objectivo de saber se os comportamentos de consumo são independentes do género e da faixa etária, recorremos a testes de independência. Em todos os testes foi considerado o nível de significância igual a 0.05.

Num segundo momento, aplicamos o modelo de regressão logística para estudar a relação entre o consumo de álcool e o insucesso escolar com base na amostra obtida.

1. Análise exploratória dos dados

Dos alunos inquiridos, 246 pertencem ao género masculino e 244 pertencem ao género feminino (correspondendo ambos a 50%, aproximadamente, da amostra).

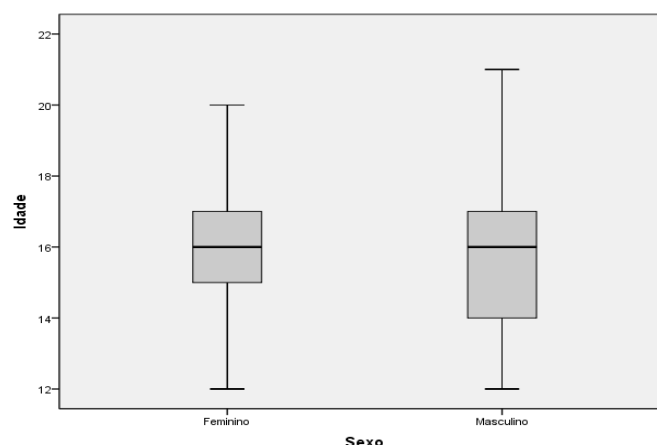


Figura 1. Variabilidade da idade segundo o género

Os inquiridos têm idade compreendida entre 12 e 21 anos, sendo a idade média igual a 15.7 (± 1.88) anos. Na amostra, e em ambos os géneros, a moda é 16 anos. Constatámos uma maior

variabilidade da idade no gênero masculino, no qual a idade média é 15.74 (± 1.96), enquanto a idade média no gênero feminino é 15.70 (± 1.82) (Figura 1).

Para testar, em cada um dos gêneros, se a variável idade segue uma distribuição normal, recorremos ao teste de ajustamento de *Kolmogorov – Smirnov* (com correção de *Lilliefors*). Os resultados dos testes (*valor - p* < 0.05 para ambos os casos) levaram a concluir que a variável idade não segue uma distribuição normal em nenhum dos gêneros.

A hipótese de que a distribuição da variável idade é a mesma nos dois gêneros foi testada através do teste não paramétrico de *Mann-Whitney*. O teste revelou evidência para afirmar que a distribuição da variável idade é igual nos dois gêneros (*valor - p* = 0.43), verificando, portanto uma igualdade entre as idades médias dos gêneros.

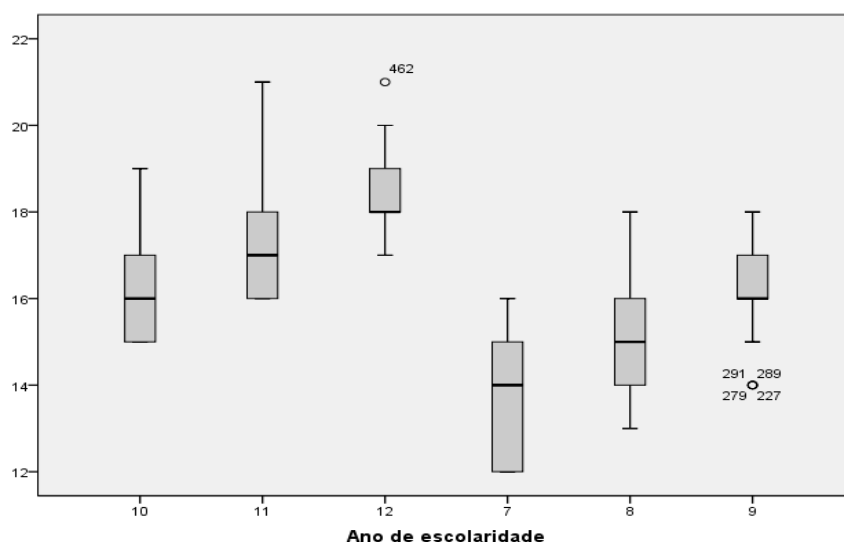


Figura 2. Variabilidade da idade segundo o ano de escolaridade

Os inquiridos distribuem-se pelos três ciclos do ensino secundário: 44.9% da amostra frequenta o 1º ciclo, 31.3% frequenta o 2º ciclo e os restantes (23.8%) frequentam o 3º ciclo. A idade dos alunos do 1º ciclo varia entre 12 e 18 anos, a dos alunos do 2º ciclo varia entre 14 e 19 anos e a idade dos alunos no último ciclo varia entre 16 e 21 anos (Figura 2). 18% dos respondentes são maiores de idade (com idade superior a 17 anos), frequentando o 9º, o 10º, o 11º ou o 12º ano de escolaridade. Atendendo à idade normal para frequência de cada ciclo, constatamos que em todos os ciclos existe uma percentagem de distorção idade-ciclo, ou seja de alunos que estão atrasados na escola.

Relativamente ao tipo de agregado familiar, a família monoparental é a mais comum entre os inquiridos (41.2%), seguida da família nuclear (35.5%). Uma percentagem relativamente baixa (2.7%) está integrada numa família reconstituída e os restantes (20.7%) referiram morar sem ambos os progenitores (sozinhos ou sob os cuidados de familiares próximos ou amigos da família).

Uma percentagem relativamente considerável dos inquiridos não apontou o nível de escolaridade e a profissão dos pais, o que, na nossa óptica, deve-se (de entre outras possíveis razões) ao facto de alguns respondentes não pertencerem ao mesmo agregado familiar dos pais. Assim, não dispomos de informação relativamente ao nível de escolaridade dos pais (21.2%) e das mães (14.9%), bem como à profissão dos pais (18.6%) e à das mães (7.3%). Devido a esta falta de informação, não utilizámos as variáveis nível de escolaridade e estatuto profissional dos pais na nossa análise. Entretanto, optamos por referir que nos dados obtidos, tanto no caso dos pais como no das mães, a maioria estudou até ao ensino básico (46.1% dos pais e 48.0% das mães) e a segunda maior porção estudou até ao ensino secundário (24.6% dos pais e 25.9% das mães). Uma percentagem dos progenitores (7.3% dos pais e 13.7% das mães) é considerada analfabeta, e ainda 22.0% dos pais e 12.5% das mães têm formação superior.

No tocante à profissão dos progenitores, constatamos que a maior porção, tanto dos pais como das mães, pertence à classe de Profissões Elementares (as mães, particularmente, são maioritariamente empregadas domésticas) e a menor porção está reformada (2.0% dos pais e 0.2% das mães). Enquanto para os pais a segunda maior porção (28.1%) pertence à classe Executante/Administração Pública, para as mães a segunda maior porção (35.2%) está desempregada. 11.0% das mães pertence à classe Executante/Administração Pública e as restantes (16.3%) pertencem à classe Independente ou à classe de Dirigentes/Técnicos, e 6.8% dos pais está desempregada e os restantes (20.0%) pertencem à classe Independente ou à classe de Dirigentes/Técnicos.

Relativamente ao percurso escolar, 26.7% dos respondentes apresenta, pelo menos, uma retenção no ensino básico e 49.2%, correspondendo à soma de 46.0% das alunas e 52.4% dos alunos, apresenta pelo menos uma reprovação no ensino secundário (Figura 3). São os alunos do 9º ano de escolaridade que apresentam o maior índice (69%) de retenções no secundário.

Os dados revelam que 42.2% da amostra (correspondendo a 85.1% dos alunos que apresentam retenções no ensino secundário) apresenta insucesso escolar em pelo menos um dos últimos

dois anos de escolaridade frequentados pelo aluno. Esta porção representa a soma de 40.6% das alunas e 44.0% dos alunos (Figura 4). Em termos percentuais, são os alunos do 11º ano que apresentam menor insucesso escolar (25.4%), enquanto os do 9º ano apresentam o maior índice de insucesso escolar (74.4%) nos dois últimos anos de escolaridade frequentados pelo aluno

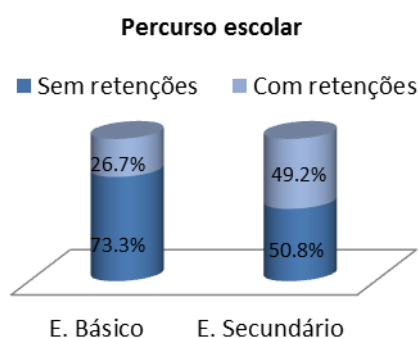


Figura 3. Percurso escolar no ensino básico e no ensino secundário

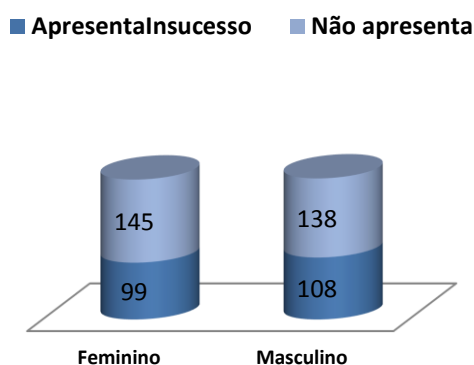


Figura 4. Insucesso escolar segundo o género

Na impossibilidade de considerar os múltiplos factores implicados no insucesso escolar, apontados na literatura, e não querendo ignorar a totalidade deles, procuramos analisar alguns itens (considerados como os mais apontados no contexto cabo-verdiano como possíveis factores associados ao (in)sucesso escolar). Assim, questionados se se consideram alunos com ou sem alguns atributos – gostar de estudar, faltar muito às aulas, dispor de bom ambiente familiar, dispor de interesse familiar nos estudos, dispor de acompanhamento dos pais/encarregado de educação nas tarefas escolares, apresentar dificuldades em acompanhar as matérias, considerar elevadas as exigências dos professores e considerar os estudos importantes para o futuro – a maioria dos inquiridos revelou considerar com os atributos que não favorecem ao insucesso escolar (Tabela 4).

Tabela 4. Insucesso escolar segundo alguns atributos

		Não apresenta insucesso		Apresenta insucesso	
		n	%	n	%
Gostar de estudar	Não	70	49.3	72	50.7
	Sim	213	61.4	134	38.6
Faltar muito às aulas	Não	247	62.7	147	37.3
	Sim	36	38.3	58	61.7
Dispor de bom ambiente familiar	Não	23	54.8	19	45.2
	Sim	260	58.2	187	41.8
Dispor de interesse familiar nos estudos	Não	53	43.1	70	56.9
	Sim	229	62.6	137	37.4
Dispor de acompanhamento dos pais nas tarefas escolares	Não	69	50.4	68	49.6
	Sim	213	60.5	139	39.5
Apresentar dificuldades em acompanhar as matérias	Não	198	62.9	117	37.1
	Sim	85	48.9	89	51.1
Considerar elevadas as exigências dos professores	Não	161	54.9	132	45.1
	Sim	122	61.9	75	38.1
Considerar os estudos importantes para o futuro	Não	10	66.7	5	33.3
	Sim	273	57.5	202	42.5

Constatámos que os alunos que referiram não gostar de estudar, os que referiram faltar muito às aulas, os que não consideram dispor de interesse familiar nos estudos, os que não consideram dispor de acompanhamento dos pais/encarregados de educação nas tarefas escolares e os que consideram apresentar dificuldades em acompanhar as matérias apresentam, em termos proporcionais, maior insucesso escolar quando comparados com os demais. Como poderemos ver mais adiante, na secção 2 deste mesmo capítulo, estas variáveis - gostar de estudar, faltar muito às aulas, interesse familiar nos estudos, acompanhamento familiar nas tarefas escolares, dificuldade em acompanhar as matérias e elevadas exigências dos professores - mostraram estar associadas ao insucesso escolar

1.1. Caracterização do comportamento de consumo de álcool

Nesta secção procedemos à análise descritiva do consumo de álcool por parte dos alunos, com base na amostra obtida, comparando os comportamentos de consumo por género e por faixa etária.

A variável idade é submetida a uma categorização assente nos pressupostos de Harper e Marshall (1991) que dividem a adolescência em três fases: a *adolescência inicial* que se situa entre os 12 e os 14 anos; a *adolescência intermédia*, entre os 15 e 17 anos; a *adolescência terminal (ou tardia)* que se situa entre os 18 e os 20 anos. Deste modo, no nosso estudo a variável faixa etária (em substituição da variável idade) recebe as seguintes categorias: *12-14 anos*, *15-17 anos* e *maiores de 17 anos*. Realçamos que, na nossa amostra encontram-se dois inquiridos com 21 anos de idade que incluímos na última fase da adolescência.

Mais de metade da amostra (57.3%) é constituída por alunos que pertencem à fase intermédia da adolescência e a menor porção (17.6%) pertence à fase terminal da adolescência (Figura 5).

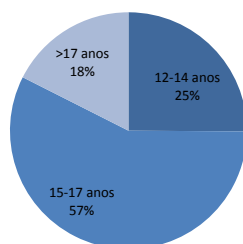


Figura 5. Distribuição da amostra por faixa etária

Ao longo desta secção a dimensão da amostra sofre alterações em função do tratamento estatístico. Em algumas situações considerámos a amostra total do estudo (n=490), noutras situações a amostra considerada refere-se apenas aos alunos que já experimentaram álcool (n=338) e, ainda, noutras situações considerámos como amostra apenas o grupo de alunos que consome álcool (n=194). Não obstante, faremos menção nos referidos quadros ou gráficos da dimensão (n) correspondente.

I. Primeira experiência com o álcool

Os dados revelam que 69.0% dos alunos já experimentou bebidas alcoólicas. Essa porção de alunos, que já tiveram uma primeira experiência com o álcool, espalha-se pelas diferentes faixas etárias e pelos dois géneros. O género feminino (71.3%) e a faixa etária dos alunos maiores de 17 anos (88.0%) constituem as maiores percentagens dos que já experimentaram álcool (Figura 6).

A primeira experiência com o álcool mostrou-se associada à faixa etária (valor- $p < 0.05$) mas independente do género (valor- $p = 0.27$). Os dados revelam que a porção de alunos que nunca experimentou uma bebida alcoólica diminui com o avançar da faixa etária.

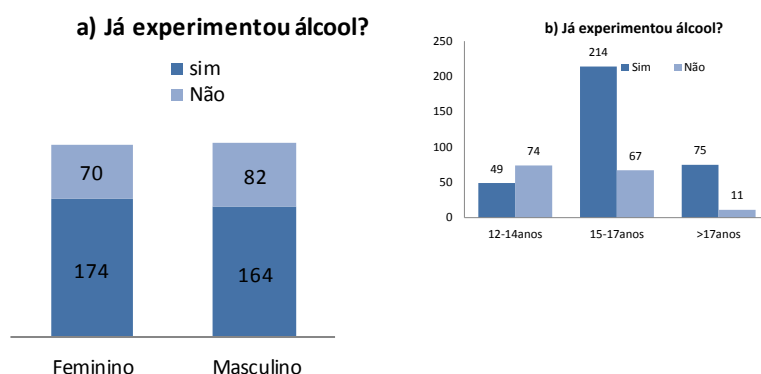


Figura 6. Primeiro consumo de álcool segundo o género e a faixa etária

A maior parte dos alunos (e seguindo a mesma tendência em ambos os géneros e nas três faixas etárias) terá tido a sua primeira ingestão de uma bebida alcoólica numa idade entre os 12 e os 17 anos, e 13.5% dos alunos terá tido esse primeiro contacto com o álcool numa idade mais precoce (antes dos 12 anos). São os rapazes e os alunos da fase inicial da adolescência que, em termos percentuais, iniciaram o consumo mais precocemente, embora a idade mais precoce (5-11 anos) tivesse sido apontada pelos alunos que estão na fase intermédia da adolescência (Tabela 5).

Tabela 5. Idade ao primeiro consumo segundo o género e a faixa etária

	Amostra (n=338)	Género		Faixa etária		
		Feminino (n=174)	Masculino (n=164)	12-14anos (n=49)	15-17anos (n=214)	>17 anos (n=75)
Idade						
0-5 anos	0.3%	0.0%	0.6%	0%	0.5%	0%
6-11anos	13.2%	10.4%	16.1%	18.4%	14.2%	6.8%
12-17anos	81.1%	85.5%	76.4%	81.6%	85.3%	68.9%
>17 anos	5.4%	4.0%	6.8%	0%	0%	24.3%

Na amostra, e em ambos os géneros, a maioria dos alunos que já experimentaram álcool elegeu uma festa como o local do primeiro consumo e os amigos como a companhia nesse consumo. Os resultados dos testes de independência evidenciaram existência de associação do local do primeiro consumo tanto com o género (valor- $p=0.05$) como com a faixa etária (valor- $p=0.03$). A análise inferencial revelou, também, evidências de associação entre a companhia no primeiro consumo e o género (valor- $p=0.02$) e associação entre a companhia no primeiro consumo e a faixa etária (valor- $p=0.03$).

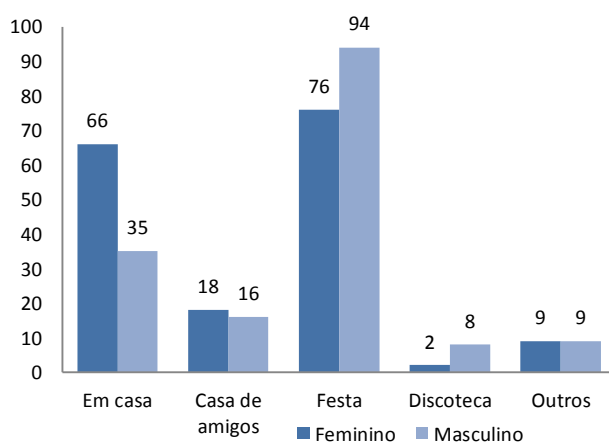


Figura 7. Local do primeiro consumo segundo o género

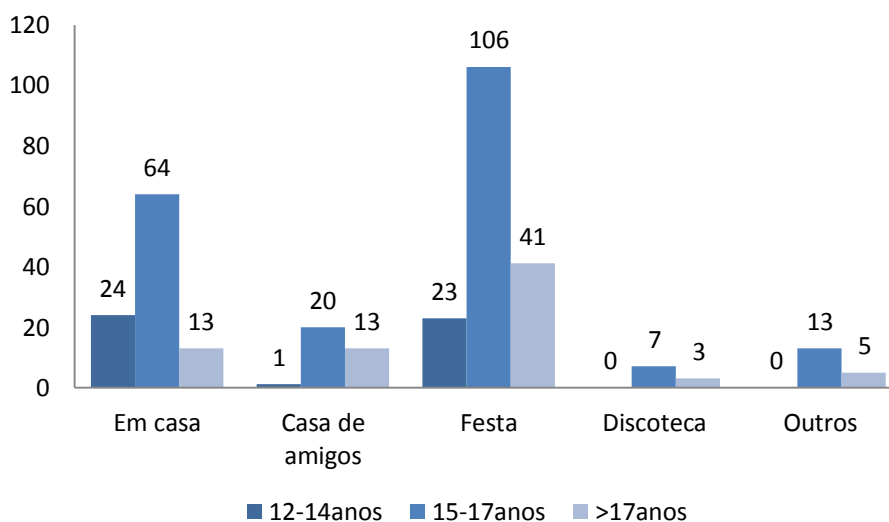


Figura 8. Local do primeiro consumo segundo a faixa etária

Relativamente ao género (pese embora os amigos fossem a companhia eleita nos dois grupos), são as alunas quem, em maior percentagem, referiram ter tido a primeira experiência com o álcool na companhia da família (38.6%) e em casa (28.8%), enquanto foram os alunos quem

mais referiram ter tido tal experiência na companhia dos amigos (74.2%) e numa festa (58.0%). No tocante à faixa etária, metade dos alunos com idade inferior a 15 anos elegeu a casa como o local da primeira experiência com o álcool (Figura 8), e são os pertencentes a essa mesma faixa etária que, em termos percentuais, mais referiram ter tido o primeiro contacto com álcool na companhia da família ou sozinhos. Constatamos que a percentagem de alunos nessas duas companhias de consumo - família e sozinho - vai diminuindo à medida que avança a faixa etária (Figura 9).



Figura 9. Companhia no primeiro consumo segundo a faixa etária e o género

II. Consumo frequente de álcool

Cerca de 40.0% da amostra, correspondendo a 57.4% dos alunos que já tiveram uma primeira experiência com o álcool, assumiu consumir álcool com alguma frequência. Desse grupo, 51.5% é do género masculino, e a maioria (62.0%) pertence à faixa intermédia da adolescência (Figura10). Em termos percentuais são os alunos com idade superior a 17 anos que mais assumiram consumir álcool (Tabela 6).



Figura 10. Distribuição do consumo por género e por faixa etária

A companhia de consumo (frequente) eleita foi a dos amigos (88.6%), sendo o consumo sozinho o menos apontado (1.6%). Verificou-se associação entre o consumo e a faixa etária (valor - $p < 0.05$) e entre a companhia de consumo e a faixa etária (valor- $p < 0.05$).

Tabela 6 . Consumo e companhia de consumo segundo o género e a faixa etária

	Amostra	Género		Faixa etária		
		Feminino	Masculino	12-14 anos	15-17 anos	>17 anos
Consumo (n=490)						
Não Consumidor	60.4%	61.5%	59.3%	84.6%	57.3%	36.0%
Consumidor	39.6%	38.5%	40.7%	15.4%	42.7%	64.0%
Companhia de consumo (n=194)						
Amigos/Colegas	88.6%	84.4%	88.9%	57.9%	87.8%	94.5%
Familiares	11.6%	14.4%	9.1%	31.6%	11.3%	5.5%
Sozinho	1.6%	1.1%	2.0%	10.5%	0.9%	0.0%

Constatamos que com o avançar da faixa etária diminui a porção de alunos abstémicos, aumenta a porção de alunos que consomem na companhia dos amigos, enquanto a porção dos que consomem na companhia dos familiares e a percentagem de alunos que consomem sozinhos diminuem (Tabela 6). É de referir ainda que nenhum aluno com idade superior a 17 anos referiu consumir álcool sozinho.

O género feminino referiu, em maior percentagem, consumir na companhia dos familiares, mas o consumo (valor- $p = 0.63$) e a companhia de consumo (valor- $p = 0.37$) mostraram-se independentes do género.

Tabela 7. Dias de semana em que ocorre o consumo segundo o género e a faixa etária

	Amostra	Género		Faixa etária		
Dias de semana		Feminino	Masculino	12-14anos	15-17anos	>17anos
	(n=194)	(n=94)	(n=100)	(n=19)	(n=120)	(n=55)
Apenas aos fins-de-semana	86.4%	89.0%	83.9%	84.2%	86.4%	87.3%
Qualquer dia de 2ª-6ªfeira	2.7%	3.3%	2.2%	0.0%	3.6%	1.8%
Qualquer dia da semana	10.9%	7.7%	14.0%	15.8%	10.0%	10.9%

Quanto aos dias da semana em que normalmente ocorre o consumo de álcool, os alunos revelaram, maioritariamente, consumir apenas aos fins-de-semana (Tabela 7). Uma percentagem de 10.9% referiu consumir em qualquer dia da semana, e são os alunos mais novos e os do género masculino, que mais apontaram essa resposta. Os dias de consumo mostraram-se independentes do género (valor- $p = 0.43$) e da faixa etária (valor- $p = 0.84$).

No atinente à frequência de consumo, a maior porção dos alunos que assumiram consumir álcool (47.4%) faz esse consumo uma vez ou menos por mês, e a porção de consumidores diminui com o aumento da periodicidade de consumo. Uma percentagem preocupante (22.4% dos consumidores ou aproximadamente 9.0% da amostra) referiu consumir álcool pelo menos uma vez por semana.

Tabela 8. Frequência de consumo segundo o género e a faixa etária

Frequência de consumo	Amostra	Género		Faixa etária		
	(n=192)	Feminino (n=91)	Masculino (n=99)	12-14anos (n=18)	15-17anos (n=119)	>17anos (n=55)
1 vez ou menos por mês	47.4%	58.1%	37.4%	52.6%	50.8%	38.2%
2-3 vezes por mês	30.2%	29.0%	31.3%	15.8%	26.3%	43.6%
Semanalmente	10.9%	6.5%	15.2%	21.1%	11.0%	7.3%
2- 3 vezes por semana	6.3%	3.2%	9.1%	10.5%	7.6%	1.8%
>3 vezes por semana	5.2%	3.2%	7.1%	0.0%	4.2%	9.1%

Os resultados dos testes de independência mostraram evidência de que a frequência de consumo está associada ao género (valor-p = 0.02). Os rapazes revelaram consumir álcool com maior frequência, e, enquanto mais de metade das alunas não abstémicas consomem álcool apenas uma vez ou menos por mês, apenas 37.4% dos alunos não abstémicos bebe álcool com esta periodicidade. Embora a frequência de consumo se tenha mostrado independente da faixa etária (valor-p = 0.17), os alunos com idade inferior a 17 anos consomem, maioritariamente, uma vez por mês, enquanto os mais velhos consomem, maioritariamente, 2-3 vezes por mês.

Os resultados chamam a atenção para o facto de a percentagem de alunos que consomem álcool uma vez ou mais por semana diminuir com o avançar da faixa etária.

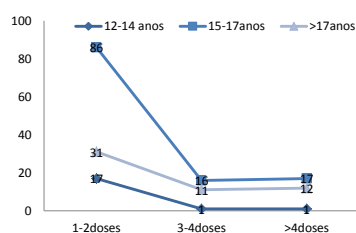


Figura 11. Quantidade consumida segundo a faixa etária

A maioria dos alunos não abstêmicos(69.8%) consome apenas 1-2 copos/doses de álcool por ocasião. São os mais velhos quem, em maior percentagem, referiu beber maiores quantidades de álcool, enquanto os mais novos (12-14 anos) revelaram beber menores quantidade de álcool por ocasião. Mas, os resultados do teste de independência (obtidos após a junção das células correspondentes às duas primeiras fases da adolescência, devido ao incumprimento dos pressupostos do teste do Qui-quadrado) mostraram evidência de que a quantidade consumida é independente da faixa etária (valor-p= 0.07).

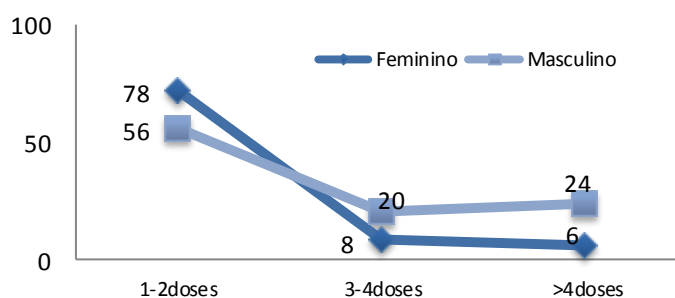


Figura 12 . Quantidade consumida segundo o género

A quantidade consumida mostrou-se estar associada ao género (valor-p<0.05). 84.8% das alunas e 56.0% dos alunos bebem, no máximo, duas doses por ocasião. Em termos percentuais, são os rapazes que ingerem maiores quantidades de álcool por ocasião (20.0% dos rapazes e 8.7% das raparigas bebem 3-4 copos/ doses por ocasião; 24.0% dos rapazes e 6.5% das raparigas bebem mais de 4 copos/doses por ocasião).

Analisando a frequência de ocorrência de consumo *binge* nos últimos 30 dias antes da aplicação do questionário, constatamos que 45.9% dos alunos não abstêmicos (correspondendo à soma de 44.7% das alunas e 47.0% dos alunos consumidores) não revelou este padrão de consumo, 23.7% revelou tê-lo tido uma vez, 20.7% teve 2 a 4 episódios e os restantes assumiram ter tido este padrão mais de 4 vezes.

São os alunos mais velhos quem mostrou ter tido o padrão de consumo *binge* com maior frequência, mas o teste de independência mostrou evidência de que a frequência de consumo *binge* é independente da faixa etária (valor-p=0.33). 10.7% dos alunos não abstêmicos, com idade inferior a 18 anos, referiu ter tido tal padrão de consumo mais de 3 vezes.

A frequência de consumo *binge* mostrou-se associada ao género (valor-p=0.002). Se por um lado, é o género masculino que, em maior percentagem, revelou nunca ter tido um episódio de

consume *binge*, por outro lado, é o mesmo gênero que (em maior percentagem) revelou ter 2 ou mais episódios desse padrão de consumo.

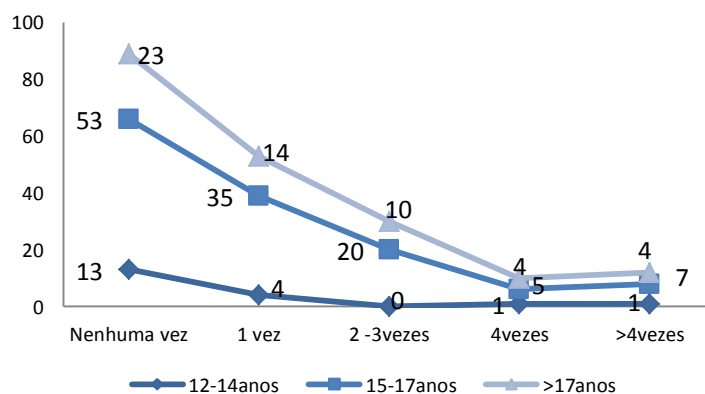


Figura 13. Consumo binge nos últimos 30 dias antes da aplicação do questionário segundo a faixa etária

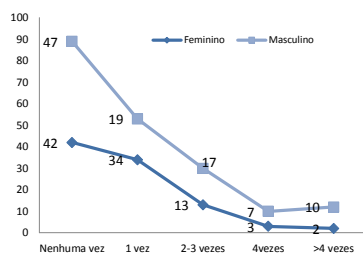


Figura 14 . Consumo binge nos últimos 30 dias antes da aplicação do questionário segundo o gênero

Os dados revelam que 25.2% dos alunos que já experimentaram álcool (correspondendo à 17.3% da amostra) já vivenciou pelo menos um episódio de embriaguez na vida. Desse grupo, a maioria (66.0%) referiu ter tido um ou dois episódios de embriaguez. Os resultados do teste de independência mostraram evidências de que existe associação entre as variáveis frequência de embriaguez e gênero (valor- $p=0.05$). É o gênero masculino que, em maior percentagem, referiu já ter-se embriagado e em maior frequência (Tabela 9).

Relativamente à faixa etária, os alunos que assumiram ter tido pelo menos um episódio de embriaguez têm idade igual ou superior a 14 anos. Os resultados do teste de independência (obtidos após a junção das células correspondentes às duas primeiras faixas etárias e da junção

das correspondentes às duas últimas categorias da variável frequência de embriaguez, devido ao incumprimento dos pressupostos do teste do Qui-quadrado) mostraram evidências de que existe associação entre a frequência de embriaguez e a faixa etária (valor- $p=0.02$). São os alunos mais velhos quem apresenta maiores frequências de embriaguez quando comparados com alunos com idade inferior a 18 anos.

Tabela 9. Frequência de embriaguez segundo o gênero e a faixa etária

	Amostra	Gênero		Faixa etária		
		Feminino	Masculino	12-14anos	15-17anos	>17anos
	(n=337)	(n=173)	(n=164)	(n=49)	(n=213)	(n=75)
Nunca	74.8%	85.0%	64.0%	91.8%	75.6%	61.3%
1-2 vezes	16.6%	13.2%	20.1%	4.1%	16.9%	24.0%
3-4 vezes	4.7%	1.2%	8.5%	4.1%	3.8%	8.0%
5-10 vezes	0.3%	0.0%	0.6%	0.0%	0.5%	0.0%
>10 vezes	3.6%	0.6%	6.7%	0.0%	3.3%	6.7%

Os dados apresentados na tabela 10 revelam que o Ponche é a bebida alcoólica mais consumida pelos alunos.

O cenário de eleição mostra diferenças em relação às bebidas alcoólicas que assumiram como a segunda opção nas faixas etárias – os alunos com idade inferior a 15 anos apresentam como a segunda opção o vinho, os na faixa 15-17 anos optam pela cerveja, enquanto os alunos mais velhos optam pelas bebidas destiladas. Os resultados do teste de independência (obtidos após a junção das células correspondentes às duas primeiras fases da adolescência e das categorias Vinho e Cerveja, devido a existência de mais de 20.0% de valores esperados inferiores a 5) revelaram existência de associação entre a faixa etária e a bebida mais consumida (valor- $p=0.02$).

Tabela 10. Bebida mais consumida segundo o gênero e a faixa etária

	Amostra	Gênero		Faixa etária		
		Feminino	Masculino	12-14anos	15-17anos	18-17anos
Bebida mais consumida	(n=194)	(n=94)	(n=100)	(n=19)	(n=120)	(n=55)
Cerveja	8.5%	5.4%	11.3%	10.5%	7.8%	9.2%
Vinho	7.9%	7.6%	8.2%	5.3%	10.3%	3.7%
Ponche	72.5%	78.0%	67.0%	73.7%	76.7%	63.0%
Bebidas destiladas (Aguardente, Vodka, Whisky, Outras)	7.9%	5.4%	10.3%	5.3%	3.4%	18.5%
Outras	3.2%	3.2%	3.0%	5.3%	1.7%	5.6%

Os resultados do teste de independência (obtidos após a junção das células relativas às categorias Vinho e Cerveja) mostraram evidências de que a bebida mais consumida é independente do género (valor- $p=0.22$).

Tabela 11. Motivações do consumo segundo o género e a faixa etária

	Amostra (n=194)	Género		Faixa etária		
		Feminino (n=94)	Masculino (n=100)	12-14 anos (n=19)	15-17 anos (n=120)	>17 anos (n=55)
Bebo para me divertir	72.2%	70.2%	74.0%	63.2%	72.5%	74.5%
Bebo porque os meus amigos bebem,	12.4%	13.8%	11.0%	10.5%	16.7%	3.6%
Bebo porque os meus pais bebem.	2.1%	0.0%	4.0%	5.3%	2.5%	0.0%
Bebo porque eu gosto.	25.8%	23.4%	28.0%	21.1%	27.5%	23.6%
Bebo para esquecer os meus problemas.	11.3%	10.6%	12.0%	21.1%	10.8%	9.1%
Outros motivos	9.8%	12.8%	7.0%	5.3%	9.2%	12.7%

No que se refere às motivações do consumo, a amostra referiu consumir bebidas alcoólicas fundamentalmente para se divertir (72.0%), sendo o consumo de álcool motivado pelo consumo paterno o menos identificado (2.1%) (Tabela 11). As motivações de consumo apresentaram-se independentes do género e da faixa etária.

A saúde (45.4%) e os estudos (23.7%) são os aspectos prejudicados pelo consumo de álcool mais referidos pela amostra, sendo a relação com os amigos o aspecto menos apontado (4.1%) (Tabela 12). Os resultados dos testes de independência revelaram evidência de associação entre o aspecto “estudos” e a faixa etária (valor- $p=0.04$), e entre o aspecto “relação com os amigos” e o género (valor- $p=0.04$).

Tabela 12. Aspectos prejudicados pelo consumo de álcool segundo o género e a faixa etária

	Amostra (n=194)	Feminino (n=94)	Masculino (n=100)	12-14anos (n=19)	15-17anos (n=120)	>17anos (n=55)
Relação com os pais	14.9%	19.1%	11.0%	15.8%	15.8%	10.9%
Relação com os amigos	4.1%	1.1%	7.0%	5.3%	5.0%	1.8%
Saúde	45.4%	51.1%	40.0%	47.4%	44.2%	47.3%
Estudos	23.7%	27.7%	20.0%	47.4%	20.8%	34.5%
Outros	14.9%	17.0%	13.0%	5.3%	15.8%	16.4%

São os alunos na fase inicial da adolescência quem, em maior percentagem, referiu que o consumo de álcool prejudica os estudos (e os alunos da fase intermédia da adolescência quem,

em menor percentagem, referiu este aspecto). Dos alunos que referiram que o álcool prejudica a relação com os amigos, a maioria (87.5%) é do género masculino.

Apenas 22.5% dos inquiridos referiu não ter amigos próximos que consomem álcool, sendo esta percentagem igual à dos alunos que referiram desconhecer quantos dos amigos próximos ingerem álcool (Tabela 13). 28.6% referiu ter poucos amigos próximos que ingerem álcool e 25.4% referiu que a maioria ou todos os amigos próximos são consumidores.

Tabela 13. Consumo familiar e consumo dos amigos próximos segundo o consumo, o género e a faixa etária

	Amostra	Consumo		Género		Faixa etária		
	(n=490)	Não consumidor (n=296)	Consumidor (n=194)	Feminino (n=244)	Masculino (n=246)	12-14 anos (n=123)	15-17 anos (n=281)	>17 anos (n=86)
Consumo por parte dos amigos mais próximos								
Nenhum	22.5%	35.3%	3.1%	21.8%	23.2%	39.8%	18.6%	10.5%
Poucos	28.6%	23.1%	37.1%	29.2%	28.0%	21.1%	30.0%	34.8%
A maioria	18.8%	11.5%	29.9%	18.9%	18.7%	4.1%	21.4%	31.4%
Todos	7.6%	2.0%	16.0%	4.9%	10.2%	4.9%	6.8%	14.0%
Não sabe	22.5%	28.1%	13.9%	25.1%	19.9%	30.1%	23.2%	9.3%
Pessoas do agregado familiar que consomem regularmente								
Pai/Padrasto	8.6%	9.1%	7.7%	6.0%	11.0%	10.6%	8.2%	7.0%
Mãe/Madrasta	2.2%	2.0%	2.6%	3.3%	1.2%	1.6%	2.5%	2.2%
Os pais	2.4%	2.7%	2.1%	2.1%	2.8%	3.3%	2.5%	1.2%
Outros (irmãos, avos, tios, outros)	13.5%	11.1%	17.0%	12.3%	14.6%	9.8%	14.2%	16.3%
Ninguém	70.0%	71.3%	68.0%	73.4%	66.7%	65.8%	70.8%	73.3%
Não responde	3.3%	3.7%	2.6%	2.9%	3.7%	8.9%	1.8%	0.0%
Pessoas do agregado familiar com problemas de alcoolismo								
Pai/Padrasto	5.1%	5.4%	4.6%	4.5%	5.7%	4.1%	5.7%	4.7%
Mãe/Madrasta	0%	0%	0%	0%	0%	0%	0%	0%
Outros (irmãos, avos, tios, outros)	14.2%	15.5%	11.9%	14.3%	13.8%	13.0%	14.5%	13.9%
Ninguém	56.7%	54.4%	60.3%	54.9%	58.5%	55.3%	55.6%	62.8%
Não sabe/Não responde	23.9%	24.7%	23.2%	26.2%	22.0%	27.6%	24.2%	18.6%

O consumo pelos amigos próximos mostrou-se associado à faixa etária (valor- $p < 0.05$) mas independente do género (valor- $p = 0.20$). Com o avançar da faixa etária aumenta a quantidade de amigos próximos que consomem e diminui a porção de alunos que desconhecem quantos dos amigos próximos ingerem bebidas alcoólicas.

Os resultados da análise estatística revelaram existência de associação entre o consumo pelos amigos próximos e o consumo (dos alunos) (valor- $p < 0.05$). Dos inquiridos que referiram não ter amigos próximos que ingerem álcool (22.5% da amostra), 94.5% referiu ser abstémico. Por outro lado, são os não abstêmicos que, em termos percentuais, apresentam mais amigos próximos que ingerem álcool.

Apenas 26.7% da amostra referiu o consumo regular na família (sendo que 3.3% não respondeu à questão relativamente ao consumo regular na família e 22.9% referiu desconhecer o consumo na família). Os progenitores ou seus cônjuges são os menos referenciados quando comparados com outros familiares (irmãos, tios, primos). O consumo regular na família mostrou-se independente do género (valor – $p = 0.11$), da faixa etária (valor- $p = 0.81$) e do consumo (dos alunos) (valor – $p = 0.47$).

A amostra referenciou uma percentagem (21.2%) de familiares com problemas de alcoolismo, sendo, igualmente, os tios, primos ou irmãos mais referenciados e a mãe nunca referenciada. O consumo problemático na família também se mostrou independente do género (valor – $p = 0.40$), da faixa etária (valor- $p = 0.62$) e do consumo (dos alunos) (valor- $p = 0.29$).

Relativamente à frequência de ocorrência de algumas consequências da ingestão do álcool, a maioria dos inquiridos referiu nunca ter sofrido nenhuma das ocorrências (Tabela 14), 6.5% referiu já ter ido a uma aula após ter bebido álcool e 9.5% referiu não se ter lembrado de nada após ter bebido álcool (oscilando as frequências entre raras vezes e muitas vezes).

Tabela 14. Frequência de ocorrência de algumas consequências do consumo de álcool

	Nunca	Raras vezes	Algumas vezes	Muitas vezes
Ir a uma aula após ter bebido álcool.	93.7%	3.3%	2.9%	0.2%
Faltar às aulas após ter bebido álcool.	98.4%	1.4%	0.2%	0.0%
Faltar às aulas por ressaca	98.2%	0.4%	1.0%	0.4%
Beber álcool antes de ir a um teste	98.0%	1.2%	0.8%	0.0%
Ter mau resultado num teste por causa das bebidas	98.6%	0.8%	0.6%	0.0%
Não se lembrar de nada após ter bebido álcool.	90.5%	5.6%	3.1%	0.8%

1.2. Descrição do insucesso escolar em função do consumo

Interessou-nos descrever o insucesso escolar em função de alguns aspectos referentes ao padrão de consumo de álcool.

Segundo os resultados, mais de metade dos alunos que consomem álcool têm insucesso escolar e à medida que aumenta a frequência de consumo aumenta a incidência de insucesso escolar, tal que: 48.4% dos alunos que consomem uma vez ou menos por mês tem insucesso escolar, dos que consomem 2-3 vezes por mês e dos que consomem mais de três vezes por mês, mais de metade apresenta insucesso escolar (Tabela 15) Segundo os resultados dos testes de independência, o insucesso escolar está associado ao consumo (valor- $p < 0.05$).

Tabela 15. Insucesso escolar segundo o consumo, a frequência de consumo e a quantidade consumida

		Não apresenta insucesso (n=283)	Apresenta insucesso (n=207)
Consumo	Não consumidor	66.2%	33.8%
	Consumidor	44.8%	55.2%
Frequência de consumo			
	1 vez ou menos por mês	51.6%	48.4%
	2-3 vezes por mês	43.1%	56.4%
	Mais de 3 vezes por mês	34.9%	65.1%
Quantidade consumida			
	1-2 copos/doses	47.8%	52.2%
	3-4 copos/doses	42.9%	57.1%
	>4 copos/doses	36.4%	63.6%

Ainda segundo os resultados apresentados na tabela 15, embora a quantidade habitualmente consumida não esteja associada ao insucesso escolar, quanto maior é a quantidade consumida maior é a incidência de insucesso escolar.

Tabela 16. Insucesso escolar segundo a frequência de consumo binge

		Não apresenta insucesso (n=283)	Apresenta insucesso (n=207)
Frequência de consumo binge			
	Nenhuma vez	52.8%	47.2%
	Uma vez	35.8%	64.2%
	2-3 vezes	56.7%	43.3%
	>3 vezes	18.2%	81.2%

Relativamente ao consumo *binge*, as maiores taxas de insucesso escolar vão para os alunos que revelaram ter tido episódios de consumo *binge* uma vez ou mais de três vezes nos últimos 30 dias antes da aplicação do questionário (Tabela 16). A frequência de consumo binge mostrou-se associada ao insucesso escolar (valor- $p = 0.01$).

2. Um modelo de regressão logística para análise da associação entre o consumo de álcool e o insucesso escolar

Nesta secção, estudamos (com base na amostra obtida) a relação entre o insucesso escolar e o consumo de álcool através do modelo de regressão logística. Relembramos que a nossa variável resposta é a variável dicotómica *insucesso escolar* (observada nos dois últimos anos de escolaridade frequentados pelo aluno) e as variáveis independentes são as apresentadas na tabela 17, sendo o consumo de álcool o factor principal em estudo. Para facilitar a nossa exposição designámos por covariáveis *tradicionais* as variáveis (apresentadas na tabela 17) numeradas de 1 a 13.

Para a análise pretendida, procurámos ajustar aos dados três modelos distintos, a saber:

- **Modelo A**, cujo objectivo é descrever o insucesso escolar em função das covariáveis *tradicionais*. Este modelo servirá de base para a comparação com o modelo final obtido.
- **Modelo B**, cujo objectivo é descrever o insucesso escolar em função do consumo de álcool (factor principal em estudo), ao lado das outras covariáveis explicativas.
- Tendo confirmado que o consumo de álcool é importante para explicar o insucesso escolar, o **Modelo C** tem como objectivo analisar o impacto da frequência de consumo no insucesso escolar, ao lado das outras covariáveis explicativas. Assim, neste último modelo o factor principal em estudo deixa de ser a variável dicotómica (consumo de álcool) passando a ser a frequência de consumo (variável número 15) que recebe quatro categorias: Não consumidor – se o aluno revelar que não consome álcool; Consumidor ocasional – se o aluno consome uma vez ou menos por mês; Consumidor menos frequente – se o aluno consome 2 ou 3 vezes por mês; Consumidor frequente – se o aluno consumir uma vez ou mais por semana.

2.1. Selecção dos modelos

Antes de ajustar cada um dos modelos referidos, procedemos a uma análise univariada, através do modelo de regressão logística, descrevendo o insucesso escolar em função de cada uma das variáveis consideradas no estudo (Tabela 17).

Tabela 17. Resultados da análise univariada – relação entre o insucesso escolar e cada uma das covariáveis

Variável /código	Categorias	Valor - p	OR	IC 95%
1. Faixa etária	12-14 anos (referência)	-	-	-
	(Id1) 15-17anos	0.000	3.50	[2.17- 5.81]
	(Id2) >17anos	0.000	4.09	[2.25- 7.58]
2. Género	Feminino (referência)			
	(Sx) Masculino	0.455	ns	ns
3. Ano de Escolaridade	7º ano(referência)			
	Ano8 8º Ano	0.021	1.91	[1.10- 3.33]
	Ano9 9º Ano	0.000	6.18	[3.39- 1.61]
	Ano10 10º Ano	0.700	1.11	[0.59 -2.08]
	Ano11 11º Ano	0.360	0.72	[0.35- 1.43]
	Ano12 12º Ano	0.739	1.12	[0.57- 2.15]
4. Agregado familiar	Nuclear (referência)			
	AgF2 Monoparental	0.690	0.92	[0.612-.39]
	AgF3 Reconstituída	0.166	0.39	[0.096- .33]
	AgF4 Outros	0.846	1.05	[0.63- 1.72]
5. Percurso básico	Sem retenções (referência)			
	Pb Com retenções	0.169	1.33	[0.89- 1.98]
6. Gostar de estudar	Não (referência)			
	Gosta Sim	0.016	0.62	[0.42- 0.91]
7. Faltar muito às aulas	Não (referência)			
	Fm Sim	0.000	2.67	[1.69- 4.27]
8. Bom ambiente familiar	Não (referência)			
	AmbF Sim	0.680	ns	ns
9. Interesse familiar nos estudos	Não (referência)	-		
	IntF Sim	0.000	0.45	[0.29- 0.68]
10. Acompanhamento familiar nas tarefas escolares	Não (referência)			
	AcomF Sim	0.040	0.66	[0.44-0.98]
11. Dificuldade em acompanhar as matérias	Não (referência)			
	Dific Sim	0.003	1.76	[1.21- 2.56]
12. Elevadas exigências dos professores	Não (referência)			
	Exig Sim	0.126	0.75	[0.52- 1.08]
13. Importância dos estudos para o futuro	Não (referência)			
	ImpE Sim	0.480	ns	ns
14. Consumo	Não Consumidor (referência)			
	Consumo Consumidor	0.000	2.41	[1.67- 3.50]
15. Frequência de Consumo	Não consome (referência)			
	Freq1 Consumo ocasional	0.007	1.92	[1.20- 3.08]
	Freq2 Consumo menos frequente	0.001	2.59	[1.46 4.62]
	Freq3 Consumo frequente	0.002	3.66	[1.89 7.32]

ns – não significativa

No que se refere às covariáveis *tradicionais*, aquelas que apresentaram associação, com valor - p inferior a 0.25 (Hosmer & Lemeshow, 1989), com o insucesso escolar foram: a faixa etária, o ano de escolaridade, o percurso básico, gostar de estudar, faltar muito às aulas, o interesse familiar nos estudos, o acompanhamento familiar nas tarefas escolares, a dificuldade em acompanhar as matérias e as exigências elevadas dos professores. Desse conjunto de variáveis candidatas a explicar o insucesso escolar, as variáveis gostar de estudar, interesse familiar nos estudos, acompanhamento familiar nas tarefas escolares e elevadas exigências dos professores apresentam-se como factores de protecção do insucesso escolar (OR “bruto” menor que 1).

Embora uma das categorias do agregado familiar tenha apresentado um valor - p significativo, optámos por não incluir essa variável no conjunto das variáveis seleccionadas para explicar o insucesso escolar uma vez que as outras categorias da mesma apresentaram valores - p muito altos. Algumas categorias da variável ano de escolaridade, também, apresentaram valores - p muito altos mas, atendendo à definição da nossa variável dependente, interessou-nos deixar esta variável no conjunto seleccionado e com a definição inicial (ou seja, sem estudar uma outra possível categorização).

Relativamente à associação principal estudada, os resultados apresentados na tabela 2 revelam que um aluno exposto ao consumo de álcool apresenta um risco 2.4 vezes maior de ter insucesso escolar quando comparado com um aluno que não está exposto ao consumo de álcool. Constata-se também que o risco de obter insucesso escolar aumenta com o aumento da frequência de consumo, oscilando esse risco entre 1.9 a 3.7 vezes superior, quando se compara um aluno consumidor com aquele que não consome.

Passaremos a descrever o processo de selecção de cada um dos modelos. Porém, antes, gostaríamos de referir que, embora a variável género não tenha apresentado associada ao insucesso escolar e estando interessados em analisar as diferenças entre os géneros, tentamos inclui-la no conjunto das variáveis candidatas para explicar a nossa variável resposta. Entretanto, o valor - p permaneceu tão alto (> 0.6) que decidimos excluir o género da análise multivariada.

I. Modelo A - modelação do insucesso escolar em função das covariáveis tradicionais

Para seleccionar o modelo que “melhor” descreve o insucesso escolar em função das covariáveis *tradicionais*, procurámos, a partir do método *stepwise* (sentido “both”), um modelo entre o modelo nulo e o modelo maximal que tenha o menor AIC (Critério de Informação de Akaike). Neste caso particular, considerámos como *modelo maximal* o modelo de regressão logística cuja variável dependente é o *insucesso escolar* e o preditor linear integra o conjunto das covariáveis *tradicionais* seleccionadas a partir da análise univariada. Recorrendo ao comando *stepAIC()* do software R, o modelo seleccionado é o apresentado de seguida, com as estimativas dos coeficientes, o teste de Wald e os respectivos níveis de significância:

Coefficients:				
	Estimativa	Desvio padrão	Teste de Wald	Valor-p
(Intercept)	-0.9621	0.3714	-2.591	p<0.01
Ano8	0.1523	0.3205	0.475	p>0.1
Ano9	0.4840	0.4047	1.196	p>0.1
Ano10	-1.5542	0.4564	-3.405	p<0.01
Ano11	-3.0085	0.6063	-4.962	p<0.01
Ano12	-3.1272	0.6717	-4.656	p<0.01
Id2	1.9120	0.3577	5.345	p<0.01
Id3	4.0589	0.6327	6.415	p<0.01
Fm1	0.6061	0.2705	2.241	p<0.05
Pb1	-0.6792	0.2987	-2.273	p<0.05
Gosta1	-0.4479	0.2439	-1.836	p>0.05
Exig1	-0.3765	0.2193	-1.717	p>0.05
Dific1	0.3472	0.2267	1.531	p>0.1

Recorrendo ao teste de Wald, ao nível de significância de 5% concluímos que as covariáveis gostar de estudar, elevadas exigências dos professores e dificuldades em acompanhar as matérias não são significativas no modelo seleccionado, pelo que foram excluídas do mesmo. Proceguindo a análise, testámos a interacção entre as variáveis que permaneceram no preditor. Apenas a interacção entre a faixa etária e o percurso básico mostrou ser significativa (valor $p=0.007$ e valor $p=0.012$ para $Id2*Pb$ e $Id3*Pb$, respectivamente).

A inclusão do termo de interacção no modelo alterou o nível de significância da variável percurso escolar (valor - $p>0.1$) pelo que testámos a hipótese de excluir essa variável (e também o termo de interacção) do modelo. Atendendo aos resultados do teste da razão de verosimilhança (valor $p=0.01$), decidimos permanecer com a covariável (e a interacção) no modelo. Assim, o **modelo A** seleccionado, que explica o insucesso escolar em função das covariáveis *tradicionais*, é definido pela expressão:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_{1i}Ano(i) + \beta_{2l}Id(l) + \beta_3Fm + \beta_4Pb + \gamma_l Id(l) * Pb$$

com as seguintes estimativas e os níveis de significância:

Coefficientes:

	Estimativa	Desvio padrão	Teste Wald	valor-p
(Intercept)	-1.6651	0.2903	-5.736	p<0.01
Ano8	0.2437	0.3162	0.771	p>0.1
Ano9	0.4436	0.4186	1.060	p>0.1
Ano10	-1.7436	0.4877	-3.575	p<0.01
Ano11	-3.1185	0.6212	-5.020	p<0.01
Ano12	-3.3718	0.6862	-4.914	p<0.01
Id2	2.4175	0.4355	5.551	p<0.01
Id3	4.7343	0.6783	6.980	p<0.01
Fm1	0.7469	0.2713	2.753	p<0.01
Pb1	0.7619	0.5244	1.453	p>0.1
Id2:Pb1	-1.7111	0.6442	-2.656	p<0.01
Id3:Pb1	-2.3004	0.9116	-2.523	p<0.05

Desvio residual: 530.93; AIC: 554.93

II. Modelo B – modelação do insucesso escolar em função da exposição ao consumo de álcool

Desta feita, o preditor linear do *modelo maximal* é definido pelo conjunto das covariáveis *tradicionais* seleccionadas a partir da análise univariada e o factor principal em estudo - consumo de álcool. O mesmo processo inicial aplicado para a selecção do *modelo A*, incluindo o teste de significância das covariáveis, permitiu-nos obter o seguinte modelo:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1Consumo + \beta_{2i}Ano(i) + \beta_{3l}Id(l) + \beta_4Fm + \beta_4Pb$$

cujas estimativas dos parâmetros e níveis de significância são apresentados de seguida:

Coefficientes:

	Estimativas	Desvio padrão	Teste Wald	Valor-p
(Intercept)	-1.4687	0.2654	-5.533	p<0.01
Consumo1	0.6774	0.2311	2.931	p<0.01
Ano8	0.1439	0.3196	0.450	p>0.1
Ano9	0.4835	0.3997	1.210	p>0.1
Ano10	-1.6973	0.4563	-3.720	p<0.01
Ano11	-2.9618	0.5941	-4.985	p<0.01
Ano12	-3.2557	0.6678	-4.875	p<0.01
Id2	1.8448	0.3576	5.160	p<0.01
Id3	3.9094	0.6326	6.180	p<0.01
Fm1	0.5709	0.2706	2.109	p<0.05
Pb1	-0.6833	0.3004	-2.274	p<0.05

Desvio residual: 530.85; AIC: 552.85

Continuando com a análise do modelo, testámos as hipóteses de interacção entre o factor exposição e cada uma das outras covariáveis. Confirmada a inexistência de qualquer interacção

com significado estatístico, as covariáveis foram testadas como potenciais factores de confusão. As alterações (superiores a 10%) observadas no efeito do factor principal sobre o insucesso escolar levaram a concluir que as covariáveis faixa etária, ano de escolaridade e faltar muito às aulas são factores de confusão.

Para controlar o efeito dos factores de confusão, recorremos ao método do *propensity score*. Definimos como *propensity score* - $ps(C_i=1/x_i)$ - a probabilidade de um aluno pertencer ao grupo de Consumidor dados os valores dos factores de confusão (faixa etária, ano de escolaridade e faltar muito às aulas). Assim, os $ps(C_i=1/x_i)$ foram estimados através de um modelo de regressão logística cuja variável dependente é o consumo de álcool e o preditor linear é definido pelo conjunto dos factores de confusão. Tendo estimado os $ps(C_i=1/x_i)$, os indivíduos foram divididos em 5 grupos de igual frequência (como sugerem Rosenbaum e Rubim, 1983), tal que para cada grupo (g_i) os indivíduos que consomem e os que não consomem álcool possuem, em média, o mesmo *propensity score*.

Prosseguindo o ajustamento do modelo, utilizámos o *propensity score* como uma variável categórica (que recebe os grupos (g_i) como as possíveis categorias), em substituição dos factores de confusão, no preditor linear. Com a inclusão do *propensity score* no modelo, a variável percurso básico deixou de ser significativa. Assim, o **modelo B** obtido é o definido pela expressão seguinte:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 \text{Consumo} + \beta_{2i} g(i)$$

com as seguintes estimativas dos parâmetros e níveis de significância

	Estimativa	Desvio padrão	Teste Wald	Valor-p
(Intercept)	-1.5598	0.2469	-6.317	p<0.01
Consumo1	0.6217	0.2137	2.909	p<0.01
g2	1.4404	0.3028	4.757	p<0.01
g3	-0.1727	0.3992	-0.433	p>0.1
g4	1.6449	0.3298	4.988	p<0.01
g5	1.4288	0.3436	4.159	p<0.01

Na tabela 18 são apresentadas as associações (*odds ratio*) entre o insucesso escolar e o consumo obtidas nos modelos multivariados ajustados - antes e depois de controlados os factores de confusão. Relembremos que o *odds ratio* obtido na análise univariada é igual a 2.41. Após o controlo dos factores de confusão, a associação entre o consumo de álcool e o insucesso escolar é traduzida por um *odds ratio* igual a 1.86.

Tabela 18. Associação entre o consumo de álcool e o insucesso escolar antes e depois da aplicação do *propensity score*

	Antes da aplicação do <i>propensity score</i>			Após a aplicação do <i>propensity score</i>		
	Valor - <i>p</i>	OR	IC 95%	Valor - <i>p</i>	OR	IC 95%
Consumo	0.003	1.97	[1.25 - 3.10]	0.004	1.86	[1.23 - 2.84]

III. Modelo C – Modelação do insucesso escolar em função da frequência de consumo ao lado das outras covariáveis explicativas

Tendo confirmado a existência de associação significativa entre o consumo de álcool e o insucesso escolar, e no intuito de averiguar se o grau desta associação varia em função da frequência de consumo, procuramos ajustar um modelo que descreve o insucesso escolar em função da frequência de consumo de álcool (ao lado das covariáveis *tradicionais*).

Desta feita, o preditor linear do *modelo maximal* é definido pelo conjunto das covariáveis *tradicionais* seleccionadas na análise univariada e a variável frequência de consumo.

Procedendo de igual modo, quanto à selecção e ao teste de significância das covariáveis aplicados para a selecção dos modelos anteriores, foi seleccionado o modelo com as seguintes estimativas e níveis de significância:

Coefficientes:

	Estimativa	Desvio padrão	Teste Wald	Valor-p
(Intercept)	-1.5067	0.2688	-5.605	p<0.01
Freq1	0.3404	0.2868	1.187	p>0.1
Freq2	0.8130	0.3528	2.305	p<0.05
Freq3	1.2582	0.3963	3.174	p<0.01
Ano8	0.2321	0.3226	0.720	p>0.1
Ano9	0.4870	0.4017	1.212	p>0.1
Ano10	-1.7393	0.4602	-3.780	p<0.01
Ano11	-2.9663	0.5975	-4.964	p<0.01
Ano12	-3.2766	0.6720	-4.876	p<0.01
Id2	1.8866	0.3587	5.259	p<0.01
Id3	3.9525	0.6374	6.201	p<0.01
Fm1	0.5602	0.2736	2.048	p<0.05
Pb1	-0.6934	0.3018	-2.297	p<0.05

Desvio Residual: 526.23; AIC: 552.23

Uma vez que a categoria Consumidor ocasional (Freq1) da variável frequência de consumo não se mostrou significativa, optámos por redefinir a variável associando as duas primeiras categorias (Não consumidor e Consumidor ocasional), passando, assim, a considerar “ Não consumidor ou Consumidor ocasional” como categoria de referência desta covariável.

Prosseguindo a análise, testámos existência de interacção entre as variáveis. Nenhuma interacção com a variável frequência de consumo mostrou-se significativa mas, existindo a

interacção significativa entre as variáveis faixa etária e percurso escolar no ensino básico, optámos por incluir esse termo de interacção no modelo. Mais uma vez a inclusão do termo de interacção alterou o nível de significância da covariável percurso básico mas, atendendo aos resultados do teste de razão de verosimilhança (valor – $p < 0.05$), permanecemos como a covariável no modelo.

Ao contrário do sucedido na selecção do *modelo B*, no *modelo C*, o factor principal (frequência de consumo) não é uma variável binária, pelo que não foi possível recorrer à aplicação do modelo da regressão logística para reduzir o viés relacionado com os factores de confusão. Procuraremos analisar tal aspecto num trabalho futuro.

Assim, o *modelo C* seleccionado é o com as estimativas e níveis de significância seguintes:

Coefficientes:

	Estimativas	Desvio padrão	Teste Wald	Valor-p
(Intercept)	-1.7414	0.2956	-5.892	$p < 0.01$
Freq2	0.6918	0.3449	2.006	$p < 0.05$
Freq3	1.1141	0.3858	2.888	$p < 0.01$
Ano8	0.2711	0.3206	0.845	$p > 0.1$
Ano9	0.3687	0.4252	0.867	$p > 0.1$
Ano10	-1.9373	0.5011	-3.866	$p < 0.01$
Ano11	-3.2151	0.6304	-5.100	$p < 0.01$
Ano12	-3.5213	0.6971	-5.051	$p < 0.01$
Id2	2.4250	0.4411	5.497	$p < 0.01$
Id3	4.6344	0.6857	6.758	$p < 0.01$
Fm1	0.6788	0.2758	2.462	$p < 0.05$
Pb1	0.6857	0.5319	1.289	$p > 0.1$
Id2:Pb1	-1.7167	0.6535	-2.627	$p < 0.01$
Id3:Pb1	-2.1597	0.9400	-2.298	$p < 0.05$

Desvio residual: 519.55; AIC: 547.55

Concluimos, assim, que a probabilidade de um aluno ter insucesso escolar é dada pela equação (com as estimativas de β_j e λ_l apresentadas acima):

$$P(\text{Insucesso}) = \frac{\exp[\beta_0 + \beta_{1(g)} \text{Freq}(g) + \beta_{2(i)} \text{Ano}(i) + \beta_{3(l)} \text{Id}(l) + \beta_4 \text{Fm} + \beta_5 \text{Pb} + \lambda_{(l)} \text{Id}(l) * \text{Pb}]}{1 + \exp[\beta_0 + \beta_{1(g)} \text{Freq}(g) + \beta_{2(i)} \text{Ano}(i) + \beta_{3(l)} \text{Id}(l) + \beta_4 \text{Fm} + \beta_5 \text{Pb} + \lambda_{(l)} \text{Id}(l) * \text{Pb}]}$$

2.2. Comparação dos modelos ajustados

Para comparar os três modelos ajustados, recorreremos à avaliação da qualidade de ajustamento, a capacidade de discriminar os indivíduos nos dois grupos da variável resposta e a interpretação dos parâmetros.

I. Qualidade de ajustamento

Para testar se cada um dos modelos se ajuste bem aos dados, aplicamos o teste de Hosmer - Lemeshow. Para os três casos, verificou-se que o teste não foi significativo (os *valores - p* são superiores a 0.05) pelo que não rejeitamos nenhuma das hipóteses nulas, ou seja, concluímos que todos os modelos ajustem-se bem aos dados.

Sendo os modelos A e C encaixados, realizámos um teste de verosimilhança para compará-los. Atendendo aos resultados (valor - $p=0.003$), concluímos que o *modelo C* ajuste-se melhor aos dados, e nota-se que este apresenta um menor AIC.

Tabela 19. Medidas de qualidade de ajustamento dos modelos

	AIC	Teste de Hosmer - Lemeshow		Curva ROC
		χ^2_{HL}	Valor - p	AUC
Modelo A	554.9	6.541	0.59	0.792
Modelo B	602.4	1.376	0.98	0.714
Modelo C	547.6	7.486	0.49	0.804

Para medir a capacidade de cada modelo classificar correctamente os alunos pertencentes à categoria “Apresenta insucesso” e os pertencentes à categoria “Não apresenta insucesso”, calculamos a área delimitada pela curva ROC (AUC) (Figura 15).

Apesar do *modelo B* revelar uma capacidade de discriminação aceitável ($AUC=0.714 > 0.6$), quando é comparado com o *modelo A*, este último apresenta uma capacidade melhor (ou seja um maior AUC - 0.792). A descrição do insucesso escolar em função da frequência de consumo, ao lado das covariáveis tradicionais, mostrou-se ser uma forma melhor de analisar o insucesso escolar em função do consumo de álcool, umavez que o valor AUC (0.804) mostra que o modelo C tem uma maior capacidade de discriminação dos grupos da variável resposta (embora não chegue a ser considerada excelente).

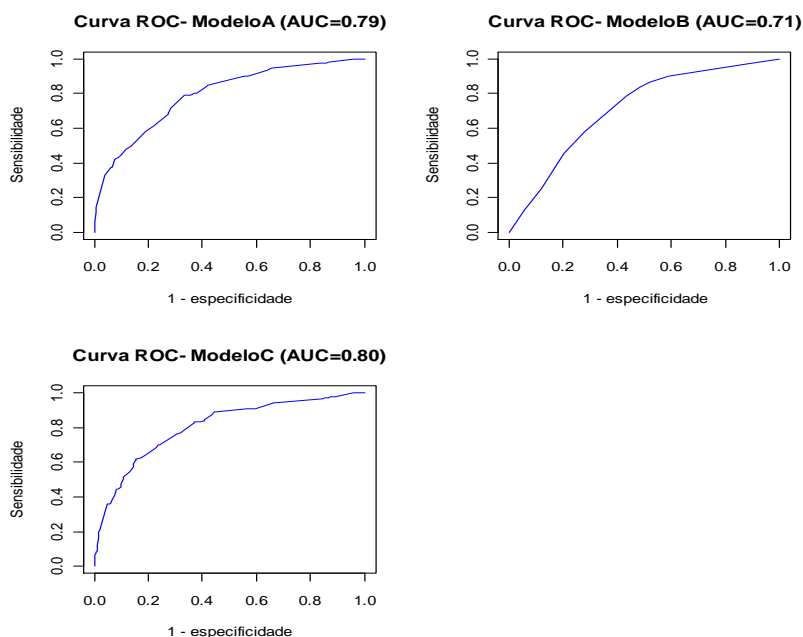


Figura 15. Curva ROC para cada um dos modelos ajustados

Portanto, o modelo C – considerado nosso modelo final no estudo - mostra ser capaz de classificar correctamente 80.4% dos alunos em todos os casos possíveis, com uma sensibilidade de 64.2% e uma especificidade de 80.8% (considerando 0.5 para o limiar).

II. Interpretação dos modelos

Embora tenhamos ajustado os dados a três modelos, a interpretação vai centrar-se no *modelo C* (modelo final) e na associação principal estudada (entre o consumo de álcool e insucesso escolar). A Tabela 5 dá-nos a estimativa das medidas de associação (*odds ratio* e intervalos de confiança) das diferentes variáveis explicativas com o insucesso escolar nos modelos A e C.

Os resultados sugerem que, embora outras variáveis se tenham mostrado (na análise preliminar) associadas ao insucesso escolar, as variáveis *tradicionais* mais importantes para explicar o insucesso escolar são a faixa etária, o ano de escolaridade, o faltar muito às aulas e o percurso escolar no ensino básico (modelos A e C). Os resultados no *modelo B* sugerem que existe uma associação significativa entre o consumo de álcool e o insucesso escolar tal que: o risco de um aluno que consome álcool ter insucesso escolar é quase o dobro (1.89 vezes maior) quando comparado com um aluno abstémico. Os resultados no *modelo C*, por sua vez, aludem

que, atender às diferentes frequências de consumo, poderá ser uma melhor forma descrever o insucesso escolar em função do consumo de álcool.

Tabela 20. Odds ratio das covariáveis e seus respectivos intervalos de confiança nos modelos A e C

Variáveis	Modelo A		Modelo C	
	OR	IC 95%	OR	IC OR
Frequência de consumo				
2-3 vezes por mês	-	-	1.99	[1.02- 3.96]
>3 vezes por mês	-	-	3.05	[1.45- 6.62]
Faixa etária				
15-17 anos	11.22	[4.89- 27.13]	11.30	[4.87- 27.64]
>17anos	113.78	[31.72- 459.89]	102.97	[28.24- 421.49]
Ano de Escolaridade				
8º ano	1.27	[0.68- 2.37]	1.31	[0.69- 2.46]
9º ano	1.56	[0.68- 3.54]	1.45	[0.62- 3.32]
10º ano	0.17	[0.07- 0.45]	0.14	[0.05- 0.38]
11ºano	0.04	[0.01- 0.14]	0.04	[0.01- 0.13]
12º ano	0.03	[0.01-0.12]	0.03	[0.01-0.11]
Faltar muito às aulas	2.11	[1.24-3.61]	1.97	[1.15- 3.40]
Percurso no ensino básico	2.14	[0.73-5.87]	1.99	[0.67- 5.51]
Interação				
(15-17anos)*P.Básico	0.18	[0.05-0.65]	0.18	[0.05- 0.66]
>17anos*P.Básico	0.10	0.016-0.57]	0.12	[0.02-0.69]

Note-se que a inclusão da variável frequência de consumo (no *modelo C*) não alterou significativamente os *odds ratio* das outras covariáveis (obtidos no *modelo A*), e a variável frequência de consumo é a segunda variável com maior associação com o insucesso escolar.

Relativamente aos resultados finais da nossa análise (obtidos no modelo C), e no que se refere às covariáveis *tradicionais*, a faixa etária, o faltar muito às aulas e o percurso no ensino básico apresentam-se como factores de risco do insucesso escolar ($OR > 1$). De acordo com os resultados obtidos, pode-se concluir que: a probabilidade de um aluno ter insucesso escolar aumenta com o avançar da faixa etária; um aluno que falta muito às aulas tem maior probabilidade (quase 2 vezes maior) de ter insucesso escolar quando comparado com aquele que não falta muito às aulas; o 10º, o 11º e o 12º ano de escolaridade apresentam-se como anos de escolaridade protectores do insucesso escolar ($OR < 1$), enquanto um aluno do 8º ano e um do 9º ano de escolaridade (apesar de apresentar maior probabilidade de insucesso escolar) não mostram diferenças significativas relativamente ao insucesso escolar, quando comparados com um aluno do 7º ano de escolaridade.

A observada interação entre as variáveis faixa etária e percurso escolar no básico sugere que: **I** - quando se comparam dois alunos que não têm retenções no ensino básico, a probabilidade de um aluno com 15-17 anos ter insucesso escolar é 11 vezes maior e a probabilidade de um aluno com mais de 17 anos ter insucesso escolar é 112 vezes maior, quando são comparados com um aluno que tem 12-14 anos; quando se comparam dois alunos que têm retenções no ensino básico, a probabilidade de um aluno com 15-17 anos ter insucesso escolar é 2 vezes maior e a do aluno com mais de 17 anos ter insucesso escolar é 11 vezes maior, quando são comparados com aquele que tem 12-14 anos; **II** - quando se comparam dois alunos que estão na primeira fase da adolescência, um aluno que tem pelo menos uma retenção no percurso básico tem maior probabilidade (2 vezes maior) de ter insucesso escolar quando comparado com aquele que não tem retenções no percurso básico; dos que estão na segunda fase da adolescência, aquele que tem pelo menos uma retenção no ensino básico tem menos chance (menos 65%) de ter insucesso escolar quando comparado com um aluno sem retenções no ensino básico; dos alunos com mais de 17 anos, aquele com pelo menos uma retenção no ensino básico tem menos 76% de chance de ter insucesso escolar, quando comparado com outro sem retenção no ensino básico.

Relativamente à associação principal em estudo, podemos concluir que a probabilidade de ter insucesso escolar aumenta com o crescer da frequência de consumo. Um aluno que consome álcool duas ou três vezes por mês tem o dobro da chance de ter insucesso escolar quando comparado com aquele que não consome ou que consome ocasionalmente. Se o aluno consumir mais de três vezes por mês a chance de ter insucesso escolar aumenta para o triplo, quando comparado com aquele que não consome ou consome ocasionalmente.

CAPÍTULO 6: DISCUSSÃO DOS RESULTADOS

Estamos cientes que os resultados referentes às outras variáveis que se mostraram associadas ao insucesso escolar merecem importantes comentários. Todavia, para não fugirmos do âmbito do nosso estudo, limitamos a discussão apenas aos resultados referentes ao comportamento dos alunos face às bebidas alcoólicas e aos resultados referentes à relação entre o consumo de álcool e o insucesso escolar.

Os resultados sugerem que quase metade dos alunos tem pelo menos uma retenção no ensino secundário. Quando nos debruçamos sobre a idade normal de frequência no ano de escolaridade, os resultados enquadram-se nos apontados pela UNICEF e ICCA (2011) que evidenciam que uma percentagem considerável de alunos estão na classe errada para a sua idade, o que pode ser consequência das retenções (ao longo do percurso escolar) e, de uma certa forma, pode influenciar o abandono do ensino secundário e, consequentemente, resultar numa menor percentagem da classe estudantil no 3º ciclo (como, também, se evidencia na amostra estudada).

Um aspecto curioso dos nossos resultados é o facto de alunos com 14 anos de idade corresponderem a *outliers* (ver Figura 2) para o 9º ano de escolaridade (quando 14 anos é idade normal para o referido ano de escolaridade). Tal facto deve-se, provavelmente, às retenções ocorridas nos primeiros anos do secundário (7º, 8º ou 9º), pois, segundo os dados do Ministério da Educação e Desporto (2011), estes são os anos com os maiores índices de retenções.

Entende-se que o insucesso escolar não é uma fatalidade mas sim um fenómeno que está relacionado com três realidades: a do aluno, a do meio social e a da escola (Benavente,1990). Atendendo aos nossos resultados o consumo de álcool faz parte da realidade dos alunos do ensino secundário e está associado ao insucesso escolar. Passaremos à discussão dos resultados referentes a esta realidade de consumo e à sua relação com o insucesso escolar.

I. Comportamentos de consumo

Enquadrando-se com os relatos em outros estudos nacionais e internacionais (Antunes & Laçador, 2011; Penchansky *et al.*, 2004; Alves, 2010), uma grande parte da nossa amostra (69.0%) referiu já ter ingerido álcool. Embora não apresentemos uma idade média para a primeira experiência com o álcool, o nosso estudo revela resultados semelhantes aos de outros estudos nacionais e internacionais, evidenciando que a maioria dos alunos teve a primeira ingestão de álcool numa idade entre 12-17 anos. Pareceu-nos mais preocupante o facto de uma percentagem de alunos (13.2%) referir ter tido o primeiro contacto com o álcool numa idade muito mais precoce (antes dos 12 anos).

Um estudo datado de 1994 indicou que, em Cabo Verde, a idade de início de consumo de álcool era 13.6 anos. Mais tarde (1998), outro estudo apontava para os 12.5 anos, e em 2007, um inquérito realizado pela Associação Zé Moniz indicou que 18.0% das crianças entre seis e 11 anos já consumia álcool. Tais resultados, e outros de estudos internacionais (Breda, 1996; Ales, 2010), evidenciam que a primeira ingestão de álcool tem ocorrido numa idade cada vez mais baixa, particularmente no género masculino. Em conformidade com o referido, os nossos dados evidenciam que são os alunos da fase inicial da adolescência (18.4%) e os rapazes que, em termos percentuais, tiveram a primeira experiência mais precocemente (16.7% dos rapazes e 10.4% das raparigas tiveram a primeira experiência antes dos 12 anos).

Freyssinet - Dominjon e Wagner (2006) referem que o início de consumo de bebidas alcoólicas é uma actividade predominantemente social, e a ligação entre o álcool e as festas é quase de ordem cerimonial, sendo o álcool, em si mesmo, o que faz de uma noite uma ocasião de festa ou de diversão. Em consonância com o sugerido pelos referidos autores, os nossos dados apontam que a maioria dos alunos experimentou o álcool numa festa e em companhia dos amigos. De facto, em Cabo Verde as festas (inclusivé as festas entre adolescentes) são, cada vez mais, ocasiões de destaque para o consumo de bebidas alcoólicas, o que revela a facilidade de acesso ao álcool e o incumprimento ou não fiscalização da lei que proíbe a venda, a oferta e o consumo de bebidas alcoólicas a menores de 18 anos.

O contexto familiar não deixa de merecer a nossa atenção, uma vez que 21.0% da amostra (contendo, particularmente, 50.0 % dos alunos com idade inferior a 15 anos) elegeu a família como a companhia e a casa como o local do primeiro consumo. Para alguns autores (Silva,

1996; Carvalho, 1997), o primeiro contacto com o álcool ocorre, em idades mais precoces, preferencialmente em casa e perante a permissividade dos pais. Em Cabo Verde, quando o primeiro consumo de álcool ocorre num contexto familiar, geralmente o consumo está associado a celebrações familiares, ao vinho e ao champanhe. Entretanto, o facto de estar a ingerir álcool em família pode dar ao adolescente a falsa noção de que este acto não acarreta consequências maléficas para a vida. A família é um ponto de referência para a vida dos adolescentes e, sendo transmissor de valores e hábitos (Cunha, 2004), a oferta de álcool ou a permissão do seu consumo (numa idade precoce) estará longe de fazer que esta seja um modelo saudável, podendo influenciar os hábitos alcoólicos dos adolescentes (Pechansky *et al.*, 2004).

Nem todos os alunos que admitiram já ter tido uma primeira experiência com o álcool, referiram consumir álcool com alguma frequência, o que poderá mostrar que, por vezes, o primeiro consumo é por mera curiosidade ou influência do momento. Os nossos resultados evidenciam uma percentagem de consumo frequente (40.0%), independente do género mas associado à faixa etária, sendo que a percentagem de alunos que consomem álcool aumenta com o avançar da faixa etária.

O consumo de álcool por adolescentes é, segundo Pechansky *et al.* (2004) e técnicos nacionais ligados à saúde e educação, o sinónimo do incumprimento e não fiscalização da lei que proíbe a venda ou oferta de bebidas alcoólicas a menores de 18 anos, da mensagem publicitária apelativa à descontração e à aventura, e que incentiva o consumo, de uma certa permissividade dos pais e da cultura da sociedade actual associada a uma mente facilmente influenciável da adolescência.

O consumo de álcool na adolescência constitui uma preocupação pelos diversos efeitos negativos que isso possa trazer, em particular, ao organismo que ainda está em desenvolvimento. A maior parte do álcool consumido é destruída no fígado, através da enzima álcool - desidrogenase. Um consumo ocasional poderá não provocar grandes transtornos, todavia, se a ingestão for frequente ou a quantidade ingerida for maior do que a capacidade do fígado, para metabolizar o álcool a concentração do álcool no organismo é maior e, piores serão as consequências para a saúde. Segundo dados científicos, 18 anos é o limiar médio de maturidade biológica no que se refere à metabolização do álcool (razão pela qual é proibida a venda, a oferta e o consumo de bebidas alcoólicas a menores de 18 anos). Assim, na adolescência (em que a enzima álcool – desidrogenase ainda existe em pouca quantidade nas

células) a degradação do álcool é muito mais difícil e os riscos para a saúde, derivados do seu consumo, são mais graves. Alguns efeitos são imediatos, mas outros são a longo prazo.

Na maioria das culturas e ao longo do tempo, seja entre os jovens, entre os adolescentes ou entre os adultos, o género masculino tem revelado consumir álcool mais do que o género feminino. No nosso estudo, embora seja o género masculino que apresenta a maior percentagem de consumo, o consumo de álcool mostrou-se independente do género. Se antes o peso cultural – o consumo de álcool pelo género feminino era pouco aceite na sociedade cabo-verdiana (e, possivelmente, noutras sociedades) – era apontado como uma das justificações para o facto de poucas raparigas admitirem consumir álcool, quando comparadas com os rapazes, hoje a fácil acessibilidade ao álcool, a maior liberdade das adolescentes e das jovens em frequentar os locais onde se consomem bebidas alcoólicas e a própria cultura social são algumas das justificações apontadas para a proximidade dos valores de consumo dos dois géneros (Carvalho, 1997; Souza *et al.*, 2005). De facto, em Cabo Verde, é fortemente visível (INE, 2008) o aumento de consumo no género feminino (e, consequentemente, esta aproximação de consumo dos dois géneros).

A maior liberdade de frequentar os locais onde se consome o álcool, bem como a maioridade e o menor controlo por parte dos pais podem ser algumas das razões pelos quais a maior incidência de consumo se regista na fase final da adolescência.

Independentemente do género, a maioria dos alunos consome na companhia dos amigos, evidenciando que o consumo de álcool é encarado como uma forma de socialização. Relativamente à faixa etária, são os alunos mais velhos quem consome mais na companhia dos amigos, e os mais novos quem consome mais entre os familiares. Mais uma vez, o contexto familiar chama-nos a atenção, evidenciando alguma permissividade de alguns pais. Se antes os pais se queixavam da má influência externa (em particular do grupo de pares) a que os filhos estão sujeitos, responsabilizando-a pelo desvio dos mesmos para o consumo de álcool, hoje, a permissividade dos pais actuais poderá ser um dos aspectos que influencia tal desvio.

Tal como é relatado noutros estudos (Freyssinet – Dominjon & Wagner, 2006; Hibell *et al.*, 2009; Alves, 2010), os alunos consomem, maioritariamente, aos fins-de-semana. Geralmente, é aos fins-de-semana que ocorrem as saídas nocturnas para locais propícios para o consumo (festas, bares e discotecas, entre outros) ou um simples encontro entre amigos, que não deixa de

ser motivo (ou uma ocasião) para consumo. Pareceu-nos preocupante o facto de 15.8% dos alunos mais novos, referirem consumir em qualquer dia da semana. Por ainda estarem sob o controlo paterno, e não tendo autorização para constantes saídas ou para frequentar locais propícios para o consumo, provavelmente, os mais novos aproveitam qualquer oportunidade (independentemente do dia da semana) para proceder ao consumo.

Relativamente à frequência de consumo e à quantidade consumida, que se apresentaram associadas ao género, os nossos resultados corroboram os de outros estudos internacionais (Alves, 2010; Freyssinet-Dominjon & Wagner, 2006), evidenciando que são os rapazes que bebem com maior frequência e em maior quantidade, quando comparados com as raparigas. Embora a frequência de consumo e a quantidade de consumo se tenham apresentado independentes da faixa etária, os nossos resultados apontam para alguns aspectos que pensamos ser importante realçar. Independentemente da faixa etária, os alunos consomem maioritariamente, 1-2 copos/doses por ocasião. Apesar de serem os alunos nas duas primeiras fases da adolescência, a referir, em maior percentagem, consumir álcool apenas uma vez ou menos por mês, a percentagem de alunos que consomem uma ou mais vezes por semana diminui com o avançar da idade. Tal resultado poderá confirmar o que têm defendido alguns agentes nacionais de saúde, referindo que, para além do início precoce de consumo, o padrão de consumo entre os mais novos revela-se cada vez mais excessivo.

Segundo alguns autores (Harnett *et al.*, 2000 e Ahlström, 2009), mesmo não consumindo com tanta frequência, é comum alguns adolescentes apresentarem uma tendência para consumir altas quantidades de álcool numa só ocasião, chegando mesmo a se embriagarem. Em consonância com o referido, os nossos resultados evidenciam uma percentagem de alunos não abstémicos (54.1%) que tiveram, pelo menos, um episódio de consumo *binge* nos últimos 30 dias antes da aplicação do questionário e outra percentagem (25.2%) que já teve pelo menos um episódio de embriaguez na vida. Independentemente da faixa etária (maior ou menor de 18 anos), os alunos que admitiram o consumo *binge* referiram, maioritariamente, ter tido apenas um episódio. À semelhança do referido por Ahlström (2009), os nossos resultados evidenciam que os rapazes apresentam episódios de consume *binge* com maiores frequências, entretanto a percentagem de alunos que teve pelo menos um episódio de consume *binge* é maior no género feminino. Tendo em conta a definição de consumo *binge* segundo o género, este resultado poderá significar que, quando o consumo é em excesso, as quantidades consumidas pelos dois géneros podem ser próximas.

Relativamente à frequência de embriaguez, os géneros e as faixas etárias (maior ou menor que 18 anos) comportam-se de formas diferentes, sendo as alunas e os alunos com idade inferior a 18 anos quem apresentam menores frequências de embriaguez.

Os comportamentos de embriaguez e de consumo *binge* devem-se muitas vezes à influência do grupo de pares. Por vezes o adolescente fala muito, e orgulhosamente, das suas experiências com o álcool aos amigos, fazendo com que estes cheguem a imaginar que o seu consumo é superior àquele que na realidade é. Esta atitude, de uma certa forma, influencia o adolescente a consumir em maior quantidade, de modo a seguir o padrão exibido pelo amigo ou pelos pares (Juniors *et al.*, 2009, citados por Alves, 2010).

Tais episódios de consumo (embriaguez e *binge*) são preocupantes, particularmente na adolescência, podendo provocar perdas de consciência (*block out*) entre outros efeitos. Os adolescentes que consomem álcool em excesso tendem a expor-se a situações de risco (como a iniciação precoce da vida sexual, gravidez precoce e experimentar outras drogas, entre outros). Aliás, no estudo de Antunes e Caçador (2012), 20.0% dos inquiridos referiram ter relações sexuais em estado de embriaguez. Por outro lado, embora não dispondo de números, alguns dos casos recentes de violência (inclusivé homicídio) ocorridos entre jovens têm sido sob efeito do álcool.

De acordo com Freyssinet-Dominjoh e Wagner (2006), as bebidas alcoólicas fazem parte de momentos de diversão. Os nossos dados vão ao encontro do sugerido pelos referidos autores na medida em que a maioria dos alunos (72.2%) referiu consumir álcool essencialmente para se divertir. De referir que uma percentagem de alunos (25.8%) consome porque gosta do álcool, ficando a ideia que alguns consomem porque têm prazer. No entanto, segundo alguns autores quando o adolescente refere gostar do álcool, isso não quer dizer que gosta do sabor, mas sim do efeito relaxante ou “desinibidor” que tal substância tem.

Ao contrário do que acontece noutros países (como Portugal, Espanha, França e Brasil, entre outros) onde a cerveja é a bebida mais consumida pelos jovens e adolescentes, a nossa amostra elegeu o Ponche como a bebida mais consumida. Este resultado vem de encontro aos resultados de Antunes e Laçador (2011). A eleição deve-se, possivelmente, ao facto de o Ponche ser (em Cabo Verde) uma das bebidas alcoólicas mais acessíveis em termos de custo e de fácil produção (sendo, muitas vezes, um produto caseiro). Se é plausível pensar que a cerveja e as

bebidas destiladas foram menos apontadas por serem ligeiramente de maiores custos, o mesmo não se pode dizer relativamente ao vinho (apontado por 7.9% da amostra). Segundo Freyssinet – Dominjon (2006) a maioria dos jovens não consome vinho porque consideram que esta bebida é para os mais adultos (pais ou avós).

Com o avançar da faixa etária, e à medida que o consumidor se habitua com o consumo de álcool, o organismo começa a exigir uma maior quantidade de álcool, para (no mínimo) manter os efeitos do consumo inicial e, conseqüentemente, as bebidas de menor teor alcoólico dão lugar às bebidas destiladas (Weeselovicz *et al.*, 2008, citado por Alves, 2010). Isso poderá explicar o facto de serem os alunos mais velhos a referir, em maior percentagem, as bebidas destiladas como as consumidas com maior frequência.

A aguardente (*grogue*) é umas das bebidas destiladas mais consumidas em Cabo Verde, inclusivé entre os mais novos. A somar ao conjunto das preocupações referentes aos aspectos que incentivam o consumo na adolescência e aos efeitos deste consumo, existe a preocupação quanto à qualidade da aguardente nacional, que é cada vez mais produzida, mais consumida e de menor custo.

Os alunos são de opinião que o álcool prejudica essencialmente a saúde e os estudos. Este resultado leva-nos a referir que deve-se repensar a informação e a educação para a saúde, junto dos nossos adolescentes, uma vez que os incentivos ou as motivações para o consumo têm falado mais alto do que o conhecimento referente aos males que isso possa trazer à vida do indivíduo.

A literatura sugere que os adolescentes são influenciados pelos seus amigos, ou tendem a relacionar-se com os de comportamento idêntico. Conseqüentemente, a maior probabilidade de um indivíduo consumir álcool ocorre entre os que se relacionam com amigos que consomem (Freyssinet-Dominjoh & Wagner, 2006; Pechansky *et al.*, 2004; Cunha & Bento, 2005). Os nossos resultados vão de encontro ao sugerido pela literatura, pois são os alunos não abstémicos, que, em maior percentagem (83%), referiram ter amigos que ingerem álcool. Relativamente à faixa etária, são os mais velhos que apresentaram mais amigos que ingerem álcool e são os mesmos que, em maior percentagem, consomem na companhia dos amigos.

Igualmente, sendo a família um agente transmissor de hábitos e valores, o hábito de consumo entre os familiares pode influenciar o adolescente a consumir álcool (Breda, 1996). Entretanto,

os nossos resultados não são suficientes para confirmar tal pressuposto, pois o consumo na família (que foi referido por uma pequena percentagem da amostra) mostrou-se independente do consumo (dos alunos), embora sejam os abstémicos que, em maior percentagem, referiram não ter familiares com hábitos de consumo de álcool. Por outro lado, e curiosamente, são os alunos que não consomem álcool quem, em maior percentagem, referiu ter problemas de alcoolismo na família.

De referir que atendendo à prevalência de consumo de álcool em Cabo Verde (53.2% da população) e às preocupações nacionais demonstradas relativamente ao aumento de consumo, a porção de alunos que referiu o consumo na família (em particular, o consumo regular) é muito baixa. Questionamos se o consumo dos pais tem sido na ausência dos adolescentes (e consequentemente estes desconhecem o consumo paterno) ou ainda se os filhos têm algum receio em referir o consumo na família.

Quisémos saber a frequência de ocorrência de algumas consequências sofridas pelo consumo de álcool mas, apesar da percentagem de alunos que consome álcool com uma certa frequência e que já teve episódios de embriaguez, quase a totalidade da amostra referiu nunca vivenciar tais consequências do consumo de álcool na escola, o que poderá ser explicado pelo facto de a maioria dos alunos consumir álcool apenas aos fins-de-semana.

II. O consumo de álcool e o insucesso escolar

Uma vez consumido, a concentração de álcool no sangue depende de factores como: quantidade de álcool consumida num determinado tempo, massa corporal, metabolismo do consumidor, quantidade de comida no estômago e particularidades do indivíduo. Geralmente, os adolescentes têm menos massa corporal do que os adultos e ainda não desenvolveram a tolerância corporal ou fisiológica ao álcool e aos seus efeitos e, portanto, não precisam consumir elevadas quantidades de álcool para sofrer fortes efeitos (Zeigler *et al.*, 2010).

Ao longo da vida o cérebro está em constante transformação, porém, as mudanças mais intensas ocorrem na adolescência (Peuker *et al.* 2006). O facto de o tecido nervoso central estar ainda em desenvolvimento, a exposição ao álcool em idades precoces poderá trazer riscos ao nível de desenvolvimento intelectual do adolescente, interferindo na capacidade de pensar, no poder de memorizar e de manter a atenção, e comprometendo as capacidades de coordenação da

informação na tomada de decisões e no controle das emoções (Dupont, 2009). A memória é a função fundamental no processo de aprendizagem e, uma vez que esta pode ser alterada com o consumo de álcool, tal consumo compromete o processo de aprendizagem e, consequentemente, o sucesso escolar.

Ao encontro destes pressupostos, os nossos resultados aludem que um aluno que consome álcool tem maior probabilidade (1.86 vezes mais) de ter insucesso escolar quando comparado com aquele que não consome. Tal resultado é consistente com alguns estudos internacionais (Alves, 2010; Peuker *et al.*, 2006; Mendes & Lopes, 2007). Alves (2010), por exemplo, conclui que um aluno que ingere álcool tem 1.88 vezes mais probabilidade de ter retenções no percurso escolar quando comparado com aquele que é abstémico.

Em consonância com o referido por Pillon e Corradi – Webster (2006), os nossos resultados revelam que os alunos que consomem álcool são os que, em maior percentagem, faltam muito às aulas, e faltar às aulas é um dos factores de risco para o insucesso escolar da população estudada.

Atendendo aos efeitos do álcool, e às especificidades da adolescência, quanto maior é a quantidade e a frequência de consumo maior será a gravidade dos efeitos na aprendizagem. Graf (2006), citado por Alves (2010), refere que a atenção, a concentração e a capacidade de raciocínio e de recolha de informações são alteradas, podendo repercutir-se no desempenho escolar. Os nossos resultados confirmam tal pressuposto, uma vez que a porção de alunos com insucesso escolar aumenta com a frequência de consumo e com a quantidade habitualmente consumida.

Peuker *et al.* (2006) referem que o padrão *binge*, em particular, é responsável por fortes alterações no desempenho de tarefas cognitivas inerentes ao funcionamento do lobo frontal associadas a quebras de tensão e perdas de memória, o que compromete o rendimento escolar. Dos nossos dados, constatamos que a incidência de insucesso escolar é maior (61.9%) no grupo de alunos que apresentaram pelo menos um episódio de consumo *binge*.

Quando o insucesso escolar é analisado em função da frequência de consumo de álcool, percebemos que não existem diferenças significativas entre os abstémicos e os que consomem ocasionalmente (uma vez ou menos por mês). Face ao resultado, Freyssinet-Dominjon e Wagner (2006) referem que a maioria dos jovens, mesmo não assumindo espontaneamente que

são consumidores, não recusam beber um copo de champanhe em determinada ocasião e tomar um aperitivo em família, mas os seus consumos são ao mesmo tempo ocasionais e irregulares, e o perfil destes está muito próximo dos abstémicos.

Relativamente às outras frequências de consumo, os nossos resultados demonstram que um aluno que consome 2 ou 3 vezes por mês tem o dobro da chance de ter insucesso escolar, quando comparado com aquele que é abstémico ou consome ocasionalmente, e o aluno que consome pelo menos uma vez por semana tem uma chance três vezes maior de ter insucesso escolar quando comparado com aquele que consome ocasionalmente ou que é abstémico. Estes resultados são consistentes com, entre outros resultados, os de Webb *et al.* (2007) que revelam que os indivíduos que apresentam retenções ingerem maiores quantidades de álcool e adoptam hábitos de ingestão mais frequentes.

Pelo exposto, percebemos que existe uma relação entre o consumo de álcool e o insucesso escolar. Segundo a literatura, tanto o insucesso escolar pode ser uma consequência do consumo de álcool como o consumo de álcool pode ser uma consequência do baixo rendimento escolar. O rendimento escolar indesejado ou aquém da real capacidade do aluno, pode fazer com que o aluno se afaste do ambiente escolar, favorecendo também o início ou a manutenção (ou mesmo o agravamento) de hábitos de consumo de álcool. Entretanto, atendendo ao nosso objectivo de chamar atenção para os prejuízos do consumo de álcool na vida do aluno/adolescente, particularmente na sua vida académica, os nossos comentários foram direccionados apenas para esse fim.

CAPÍTULO 7: CONCLUSÃO

Neste estudo, como já foi referido, quisemos conhecer o padrão de consumo entre os alunos do ensino secundário da ilha de São Vicente (Cabo Verde) e estudar a sua relação com o insucesso escolar.

Concluimos que mais de metade dos alunos já experimentou álcool e, independentemente do género, o primeiro consumo de álcool ocorre maioritariamente entre os 12 e os 17 anos. Existe uma percentagem de alunos, particularmente entre os mais novos (18.4%), que tiveram a primeira experiência com o álcool numa idade muito mais precoce (antes dos 12 anos), o que leva a concluir que o início do consumo tende a ocorrer numa idade cada vez mais baixa. Este primeiro contacto com o álcool ocorreu, maioritariamente, numa festa e na companhia dos amigos, e uma considerável percentagem fê-lo num ambiente familiar (em casa ou na companhia de familiares).

Concluimos que o consumo de álcool entre os alunos das escolas secundárias de São Vicente (Cabo Verde) é uma realidade, sendo que 40.0% dos alunos consomem álcool. Independentemente do género e da faixa etária, os alunos consomem, maioritariamente, aos fins - de - semana e na companhia dos amigos. São os mais velhos que mais consomem, mas a frequência de consumo e a quantidade consumida levam a alguma preocupação, uma vez que os alunos com idade inferior a 18 anos tendem a seguir um padrão semelhante aos de idade superior a 17 anos. Relativamente ao género, são os rapazes que consomem com maior frequência e em maiores quantidades.

A facilidade de acesso, a produção e o baixo custo devem ser algumas das justificações do Ponche ser a bebida alcoólica mais consumida entre os alunos. Existe, porém, uma percentagem, fundamentalmente de alunos mais velhos, que consome bebidas destiladas com maior frequência.

Pensamos ser preocupante o facto de um pouco mais de metade dos alunos não abstémicos (correspondendo a 21.4% da amostra estudada) ter tido, pelo menos, um episódio de consumo

binge nos últimos 30 dias antes da aplicação do questionário, e 25.0% da amostra já ter pelo menos um episódio de embriaguez na vida.

Concluimos que são os alunos que consomem álcool e os que apresentam episódios de consumo *binge* os que apresentam maiores índices insucesso escolar quando comparados com os demais, e à medida que aumenta a frequência de consumo e a quantidade consumida aumenta a incidência de insucesso escolar.

O consumo de álcool constitui um factor de risco para o insucesso escolar dos alunos do ensino secundário, tal que a probabilidade de um aluno não abstémico ter insucesso escolar é 1.86 vezes maior quando comparado com aquele que é abstémico. Concluimos ainda que o grau de associação (*odds ratio*) varia em função da frequência de consumo, isto é, quanto maior é a frequência de consumo maior é a probabilidade de um aluno ter insucesso escolar.

Não ignoramos o facto de, assim como o álcool poder ser um dos responsáveis do insucesso escolar entre os nossos alunos, o baixo rendimento escolar poderá também favorecer o início ou a persistência dos hábitos de consumo de álcool. Diante dos nossos resultados sugerimos que, não ignorando que o consumo de álcool é uma realidade entre os alunos do ensino secundário e que tal consumo tem graves efeitos para a vida dos mesmos, os agentes da educação, os pais/encarregados de educação e toda a comunidade educativa atendam à relação existente entre os dois fenómenos – consumo de álcool e insucesso escolar - agindo para que, combatendo um destes fenómenos poderá estar a combater ou a minimizar o outro.

Pensamos que uma forma de minimizar tal fenómeno é através da informação e da educação de modo que os alunos (e adolescentes) não sejam levados pela falsa imagem (de sucesso e de alegria, entre outros) que a publicidade e a própria cultura social têm dado ao álcool, ocultando os males que isso possa trazer à vida do indivíduo, particularmente à de um adolescente. Por outro lado, medidas de ordem legislativa que fizessem cumprir a lei que proíbe a venda e a oferta de bebidas alcoólicas a menores de 18 anos, também ajudariam a minimizar a facilidade de acesso ao álcool.

É na escola onde os adolescentes passam a maior parte do tempo. Sem minimizar a responsabilidade dos pais (que aliás, pensamos ser crucial), a escola pode ser um espaço de investimento tanto da educação como da saúde dos seus alunos proporcionando actividades no

sentido de prevenção de consumo de álcool. A educação para a saúde dos adolescentes, em particular os seus hábitos de consumo, também está associada ao grau de permissividade dos pais. Relembramos que uma percentagem da nossa amostra referiu ter o primeiro contacto com o álcool na companhia da família, o que demonstra que o combate ao consumo de álcool na adolescência (e ao insucesso escolar) requer, também, um trabalho junto dos pais/encarregados de educação. Aliás, o interesse familiar, que no nosso país se traduz numa relação entre a escola e o encarregado de educação, é (segundo os nossos resultados) um dos factores protectores do insucesso escolar.

Portanto, pensamos que uma intervenção em parceria – escola, ministério da saúde, pais/encarregados de educação e autoridades - seria uma estratégia de promoção da saúde dos nossos alunos e adolescentes.

Os resultados mostram, como sugere Breda (1996), que o insucesso escolar depende, também, da realidade do educando. Insistentemente, a escola tem apostado em actividades como repetição de avaliações (por vezes, correndo o risco de baixar o nível de exigência), aulas e provas de recuperação, no sentido de combater o insucesso escolar (que tem sido medido apenas através das classificações/avaliações nas disciplinas), ignorando, na prática, a realidade do aluno que vai além do aspecto cognitivo. Acreditando que uma *escola para todos* é aquela capaz de atender às necessidades (ou particularidades) do universo estudantil de modo a minimizar as diferenças entre as condições de possibilidade de sucesso, pensamos ser importante atender a que o consumo de álcool (motivado por factores diversos) também faz parte da realidade dos educandos de hoje, e esta (ou qualquer outra realidade) sempre que possível deverá ser tida em conta no empenho do desenvolvimento e da educação dos nossos educandos.

Finalmente, pensamos ser importante realçar que a realidade que conhecemos (como agentes da educação) bem como os resultados do nosso estudo, nos levam a referir que, se é certo que a maioria dos nossos alunos não consome álcool ou consome ocasionalmente, também é certo que existe uma percentagem preocupante que consome e o trabalho que se mostra ser necessário fazer deverá ser dirigido aos que consomem (no sentido de combate) bem com aos que não consomem (no sentido de prevenção).

BIBLIOGRAFIA

Afonso, M. M. (2002). *Educação e Classes Sociais em Cabo Verde*. Lisboa. Spleen Edições.

Ahlström, S. (2009). Consumo nocivo de álcool entre estudantes europeus: resultados do ESPAD. In Andrade, A. G., Anthony, J.C.& Silveira, C. M. *Álcool e suas Consequências: uma Abordagem Multiconceitual*. (Capítulo 4). São Paulo: Editora Manole.

Alves, A. L. V. (2010). *Consumo de Bebidas Alcoólicas, Padrão “Binge Drinking” e sua Relação com o Insucesso Escolar*. Tese de Doutorado em Ciências da Educação. Universidade Trás-os Montes e Alto Douro.

Alves, R. & Kossobudzky, L.A (2002). Caracterização dos adolescentes internados por álcool e outras drogas na cidade de Curitiba. *Interação em Psicologia*, 6(1), 65-79.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Auto. Cntl., AC 19*, 716-723.

Andrade, A. G. & Oliveira, L. G., (2009). Principais consequências em longo prazo relacionadas ao consumo moderado de álcool. In Andrade, A. G., Anthony, J.C.& Silveira, C. M. *Álcool e suas Consequências: uma Abordagem Multiconceitual* (Capítulo 3).São Paulo: Editora Manole.

Anguita, J. C. & Lizana, J.P.L. (1996). Padron de consumo juvenil de alcohol y problemas asociados. *Medicina Clinica*, 107 (14), 544-548.

Antunes, S. & Caçador, J. (2012). *Adolescência, Álcool e Comportamentos de Risco* (São Vicente - Cabo Verde). Relatório de Estágio. Escola Superior de Enfermagem de Lisboa.

Aragão, M. J. & Sacadura, R. (2002) *Guia Geral das Drogas: Explicar o seu Mecanismo e as suas Consequências*. Lisboa: Terramar.

- Babor, T. F., Fuente, J. R., Saunders, U. & Grant, M. (1992). AUDIT. The alcohol use disorders identification test: guidelines for use in primary health care. *WHO/PAHO*, 4, 1 – 29.
- Benavente, A. (1990). Insucesso escolar no contexto português – abordagens, concepções e políticas. *Análise Social*, 25 (108), 715-733.
- Breda, J. (1996). Bebidas alcoólicas e jovens – Um estudo sobre consumos, conhecimentos e atitudes. *Revista da Sociedade Portuguesa de Alcoologia*, 6(3), 37-52.
- Carvalho, A. A. (2004b). Políticas públicas sobre álcool. In Ferreira – Borges, C e Filho, H. C (Coord): *Alcoolismo e Toxicodependência: Manual Técnico* 2. (pp 163-177). Lisboa: Climepsi Editores.
- Carvalho A., Lemos E., Raimundo, F., Costa, M. & Cardoso, F. (2007). Caracterização de consumo de substâncias psicoactivas numa população escolar. *Toxicodependências*, 13(3), 31-36.
- D'Agostinho, R.B. Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17 (19), 2265-2281.
- Facci, M. G. D., Tessaro, N. S., Leal, Z. F. R. G., Silva, V. G. & Roma, C.G. (2007). Psicologia histórico-cultural e avaliação psicológica: o processo ensino aprendizagem em questão. *Revista Semestral da Associação Brasileira de Psicologia Escolar e Educacional*, 11(2), 323-338.
- Fahrmer, L. & Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, 13, 342-368.
- Ferreira, M. J. & Campos, P. (2001). *O inquérito estatístico: uma introdução à elaboração de questionários, amostragem, organização e apresentação dos resultados*. Disponível em: http://alea.ine.pt/html/dossier/html/meio_dossier11.html.
- Freyssinet – Dominjon, J. & Wagner, A. C. (2006). *Os Estudantes e o Álcool*. Coimbra: Quarteto Editora.

- Formosinho, J. (1991). A igualdade em educação. In *A Construção Social da Educação Escolar*, Col. Biblioteca básica de educação e ensino. Edições ASA/Clube de professores. Rio Tinto, 169-186.
- Gaspar, T., Gonçalves A., Ramos, V. & Matos, M. G. (2008). Desvantagem socioeconómica, etnicidade e consumo de álcool na adolescência. *Análise Psicológica*, 4(24), 495-508.
- Harnett, R. Thom, B., Herring, R. & Kelly, M. (2000). Alcohol in transition: towards a model of young men's drinking styles. *Journal of Youth Studies*, 3(1), 61-77.
- Harper, J. & Marshall, E. (1991). Adolescents problems and their relationship so self-esteem. *Adolescence*, 26 (104),799-808.
- Hil, M. M. & Hill, A. (2005). *Investigação por Questionário*. Lisboa. Edições Silabo.
- Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley, New York.
- Instituto Nacional de Estatística (INE), Ministério da Saúde & Organização da Nações Unidas (ONU). (2008). *Mulheres e Homens em Cabo Verde – IDNT, 2007*. Praia – Cabo Verde.
- Maroco, J. (2003). *Análise Estatística com Utilização do SPSS*. Edições Silabo.
- Marinho, R. T. (2005). Álcool e sinistralidade rodoviária. *Revista Ordem dos Médicos*, 21 (56), 10 – 12.
- Martin, C. S. & Winters, K. C. (1998). Diagnosis and assessment of alcohol use disorders among adolescents. *Alcohol Health & Research World*, 22:95-106.
- Mendes, V. & Lopes, P. (2007). Hábitos de consumo de álcool em adolescentes. *Toxicodependências*, 13(2),24-40.
- McCullagh, P. & Nelder, J.A. (1989). *Generalized linear models*. 2nd edition. Chapman and Hall, London.
- Ministério da Educação e Desporto (2011). *Os Principais Indicadores da Educação 2010/2011*. Praia – Cabo Verde.

Ministério da Saúde de Cabo Verde (2008). *Plano Nacional de Desenvolvimento Sanitário 2008-2011*, Vol. 2. Praia – Cabo Verde.

Moreira, P. (2002). *Para uma Prevenção que Previna*. Coimbra. Quarteto Editora.

Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A* 135, 370-384.

Oliveira, G. A. (2009). *Bioestatística, Epidemiologia e Investigação*. Lisboa. LIDEL.

Pechansky, F. & Barros F. (1995). Problems related to alcohol consumption by adolescents living in the city of Porto Alegre, Brazil. *Journal of Drug Issues*, 25(4), 735-50.

Pechansky, F., Szobot, C. M. & Scivoletto, S. (2004). Uso de álcool entre adolescentes: conceitos, características epidemiológicas e factores etiopatogénicos. *Revista Brasileira de Psiquiatria*, 26 (1), 14-17.

Peuker A. C., Fogaça, J. & Bizarro, L. (2006). Expectativas e beber problemático entre universitários. *Psicologia: Teoria e Pesquisa*, 22(2), 193-200.

Pillon, S.C. & Corradi – Webster, C. M. (2006). Teste de identificação de problemas relacionados ao uso de álcool entre estudantes universitários. *Revista de Enfermagem UERJ*, 14(3), 325-332.

Ribeiro C (2008). Álcool – Impacto no indivíduo e na sociedade. Qual o papel dos Cuidados de Saúde Primários? *Revista Portuguesa Clínica Geral*, 25, 269-274.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41 – 55.

Ryback, R S. (1971). The continuum and specificity of the effects of alcohol on memory. *Quart. J. Stud Alcohol*, 32, 995-1016.

Sekhon, J.S., (2011). Multivariate and propensity score matching software with automated balance optimization: The matching package for R. *Journal of Statistical Software*, Vol. 42.

Silva, A. B., Matos, M. M. G. & Diniz, J. A. (2008). Consumo de substâncias e satisfação com a vida nos adolescentes. In Matos, M.G. (Coord). *Consumo de Substâncias. Estilo de Vida? À Procura de um Estilo?* (pp.71-94). Lisboa: Instituto da Droga e da Toxicodependência.

Spazziani, M. L. (2001). A saúde na escola: da medicalização à perspectiva da psicologia histórico-cultural. *ETD - Educação Temática Digital*, 3 (1), 41-62.

Turkman, M. A. A. & Silva, G.L. (2000). *Modelos Lineares Generalizados – da Teoria à Prática*. Edições SPE. Lisboa.

UNICEF & Instituto Caboverdiano da Criança e da Adolescente (ICCA) (2011). *Análise de Situação da Criança e da Adolescente em Cabo Verde*. Praia – Cabo Verde.

Webb, J. A., Moore, T., Rhatigan, D., Stewart, C. & Getz, G. (2007). Gender differences in the mediated relationship between alcohol use and academic motivations among late adolescents. *American Journal of Orthopsychiatry*, 77 (3), 478-488.

Williams, D.A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, 36, 181-191.

WHO Constitution. (1948). Disponível em: <http://www.who.int/governance/eb/constitution/en/index.html>. Acesso em 23 de Março de 2012.

Zeigler, D.W., Wang, C.C., Yoast, R. A., Dickinson, B. D., MCaffree, M.A., Robinowitz, C.B. & Sterlina, M.L. (2005). The neurocognitive effects of alcohol on adolescents and college students. *Preventive Medicine*, 40, 23-32.

ANEXOS

Anexo A – Autorização para aplicação do questionário

Autorizado
09/02/12



Exmo. Sr. Diretor da
Escola Secundária José Augusto Pinto
São Vicente – Cabo Verde

Lisboa, 05 de Fevereiro 2012

Assunto: Pedido de autorização para aplicação de um questionário no âmbito de um Mestrado em Bioestatística.

Eu, Crisolita Sousa de Brito, professora de Matemática do ensino secundário em Cabo Verde, em comissão eventual de serviço, e aluna do Mestrado em Bioestatística na Faculdade de Ciências da Universidade de Lisboa, pretendo desenvolver um estudo intitulado “A influência do consumo de álcool no insucesso escolar dos alunos das escolas secundárias de São Vicente – Cabo Verde” orientado pela docente da referida Faculdade, Prof. Fernanda Diamantino e que culminará com a dissertação do Mestrado. A amostra do referido estudo será constituída por alunos de todas as escolas secundárias públicas de São Vicente - Cabo Verde.

É neste contexto que venho, por este meio, solicitar a V. Ex.^a a sua autorização para aplicar um inquérito, na segunda semana de Março, aos alunos desta escola. Mais adiante que toda a informação que venha a ser recolhida será tratada com anonimato, e que estou ao seu dispor para esclarecimento de qualquer informação que julgar oportuno.

Agradecendo desde já a sua colaboração, os meus melhores cumprimentos.

Atenciosamente,

Crisolita Sousa de Brito

(Crisolita Sousa de Brito)

Anexo B – O questionário

Este questionário faz parte de um estudo, intitulado “A influência do consumo de álcool no insucesso escolar dos alunos do ensino secundário de São Vicente – Cabo Verde”, para a obtenção do grau de mestre em Bioestatística. Sendo ANÓNIMO, gostaria que respondesses a cada uma das questões que se segue, com espontaneidade e sinceridade, assinalando sempre com um X no quadrado ou espaço correspondente à tua opção.

Idade _____

Sexo: ☐ F ☐ M

Ano Escolar: ☐ 7º ☐ 8º ☐ 9º ☐ 10º ☐ 11º ☐ 12º

I. INFORMAÇÃO FAMILIAR

1. Com quem vives? (escolha múltipla)

☐ Pai ☐ Mãe ☐ Irmãos ☐ Avós ☐ Outros (Quem?) _____

2. Qual o nível de escolaridade dos teus pais?

- Pai: ☐ Analfabeto ☐ 0-6º Ano ☐ 7º-10ºano ☐ 11º-12ºano ☐ Formação Superior
- Mãe: ☐ Analfabeta ☐ 0-6º Ano ☐ 7º-10ºano ☐ 11º-12ºano ☐ Formação Superior

3. Qual a profissão dos teus pais?

Pai: _____ Mãe: _____

II. PERCURSO ESCOLAR

1. Já alguma vez repetiste um ano escolar? ☐ Sim ☐ Não

1.1. Se “Sim”, assinala no ano de reprovação o número de vezes que reprovaste:

Ano de escolaridade	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º	11º	12º
Número de reprovações												

2. Assinala “Sim” ou “Não” como resposta, se te consideras um aluno com cada um dos atributos abaixo indicados.

Atributos	Sim	Não
Gosta de estudar.		
Falta muito às aulas.		
Tem um bom ambiente familiar.		
Dispõe de interesse familiar para os estudos.		
Dispõe de acompanhamento familiar (ou encarregado de educação) na orientação dos estudos.		
Tem dificuldade em acompanhar as matérias.		
Considera elevadas as exigências dos professores.		
Considera os estudos importantes para o seu futuro.		

III. COMPORTAMENTO EM RELAÇÃO AO CONSUMO DE ÁLCOOL

-
1. **Alguma vez já bebeste/provaste uma bebida alcoólica?** ☐ Sim ☐ Não
 2. **Com que idade bebeste/provaste álcool, pela primeira vez?**
☐ Nunca bebi ☐ 0 - 5anos ☐ 6 - 11anos ☐ 12 -17anos ☐ Com mais de 17anos
 3. **Na primeira vez que bebeste álcool, com quem estavas?**
☐ Nunca bebi ☐ Sozinho ☐ Com a família ☐ Com amigos ☐ Outro(Quem?)_____
 4. **Na primeira vez que bebeste álcool, onde estavas?**
☐ Nunca bebi ☐ Em casa ☐ Em casa de um amigo
☐ Numa festa ☐ Numa discoteca ☐ Outro (Onde?) _____
 5. **Actualmente, bebes bebidas alcoólicas?** ☐ Sim ☐ Não
 6. **Com que idade começaste a beber regularmente?** _____
 7. **Actualmente, bebes na companhia de quem?**
☐ Não bebo ☐ Amigos ou Colegas ☐ Familiares ☐ Sozinho
 8. **Em que dias da semana costumas beber álcool?**
☐ Não bebo ☐ Apenas aos fins de semana ☐ Qualquer dia entre 2ª e 6ª feira
☐ Qualquer dia da semana
 9. **Com que frequência costumas beber bebidas alcoólicas?**
☐ Não bebo ☐ Uma vez ou menos por mês ☐ De 2 a 3 vezes por mês
☐ Semanalmente ☐ 2 a 3 vezes por semana ☐ Mais do que 3 vezes por semana
 10. **Quando bebes álcool, quantos copos/doses costumas beber (no dia)?**
☐ Não bebo ☐ 1 a 2 copos/doses ☐ De 3 a 4 copos/doses ☐ Mais de 4 copos/doses
 11. **Responde a esta questão, atendendo ao teu sexo.**
Masculino: Nos últimos 30 dias, com que frequência bebeste mais de 4 doses de álcool numa única ocasião?
☐ Nenhuma vez ☐ 1 vez ☐ 2 a 3 vezes ☐ 4 vezes ☐ Mais do que 4 vezes
Feminino: Nos últimos 30 dias com que frequência bebeste mais de 3 doses de álcool numa única ocasião?
☐ Nenhuma vez ☐ 1 vez ☐ 2 a 3 vezes ☐ 4 vezes ☐ Mais do que 4 vezes
 12. **Alguma vez já te embriagaste?** ☐ Nunca ☐ 1 a 2 vezes ☐ 3 a 4 vezes
☐ 5 a 10 vezes ☐ Mais de 10 vezes
 13. **Qual das bebidas alcoólicas bebes com maior frequência?**
☐ Não bebo ☐ Cerveja ☐ Vinho ☐ Ponche
☐ Bebidas destiladas (aguardente, whisky, vodka, outras) ☐ Outro. Qual? _____
 14. **Que razões te levam a beber álcool? (escolha múltipla)**
☐ Para me divertir ☐ Porque os meus amigos bebem ☐ Porque os meus pais bebem
☐ Porque gosto ☐ Para esquecer os meus problemas ☐ Outras(Quais?) _____
-

15. O fato de beberes bebidas alcoólicas prejudica-te em que aspetos? (escolha múltipla)

- ☐ Não bebo ☐ Relação com os pais ☐ Relação com os amigos ☐ Saúde
☐ Estudos ☐ Outro (Qual?) _____

16. Entre os teus amigos mais próximos, quantos costumam beber?

- ☐ Nenhum ☐ Poucos ☐ A maioria ☐ Todos ☐ Não sei

17. Em tua casa, alguém bebe álcool regularmente?

- ☐ Sim. Quem? (Grau de parentesco) _____ ☐ Não

18. Na tua opinião, alguém da tua família (agregado familiar) tem problemas de alcoolismo?

- ☐ Sim. Quem (Grau de parentesco)? _____ ☐ Não ☐ Não sei

19. Assinala, para cada afirmação, a opção que mais se aproxima da frequência com que cada uma das situações te aconteceu.

Afirmação	Nunca	Raras vezes	Algumas vezes	Muitas vezes
Ir a uma aula após ter bebido álcool.				
Faltar às aulas após ter bebido álcool.				
Faltar às aulas por ressaca.				
Beber álcool antes de ir a um teste.				
Ter mau resultado num teste por causa das bebidas.				
Não se lembrar de nada após ter bebido álcool.				

Agradeço a tua preciosa colaboração, pois sem ela não é possível a realização do estudo!
 Crisolita Sousa de Brito

Anexo C – Código do R

Apresentam-se apenas os códigos do R utilizados na construção dos modelos (A, B e C); dados é a matriz que contém os dados (as observações)

1. Construção do Modelo A

```
>modelo0=glm(Insucesso~1,family=binomial("logit"), data=dados) # modelo
## nulo
>library (MASS)

## O stepAIC selecciona o modelo, entre o modelo nulo e o modelo máximo, que
tenha o menor AIC;
>stepA=stepAIC(modelo0,direction="both",scope=list(upper=~Id+Ano+Pb+Gost
a+Fm+I ntF+AcompF+Dific+Exig, lower0~1),trace=1)
>stepA$anova

## Analisando a significância dos parâmetros no modelo seleccionado pelo
stepwise
>summary(stepA)
>modA=glm(insucesso ~ Ano + Id + Fm + Pb, family = binomial("logit"), data
= dados)
## Testando a interacção entre variáveis:
>modA1=glm(Insucesso ~ Ano + Id + Fm + Pb+ Ano*Id, family =
binomial("logit"), data = dados)
>modA2=glm(Insucesso ~ Ano + Id + Fm + Pb+ Ano*Fm, family =
binomial("logit"), data = dados)
>modA3=glm(Insucesso ~ Ano + Id + Fm + Pb+ Ano*Pb, family =
binomial("logit"), data = dados)
>modA4=glm(Insucesso ~ Ano + Id + Fm + Pb+ Id*Fm, family =
binomial("logit"), data = dados)
>modA5=glm(Insucesso ~ Ano + Id + Fm + Pb+ Fm*Pb, family =
binomial("logit"), data = dados)
>summary(modA1)
>summary(modA2)
>summary(modA3)
>summary(modA4)
>summary(modA5)

## Modelo A seleccionado:
>modeloA=glm(Insucesso ~ Ano + Id + Fm + Pb + Id * Pb,
family = binomial("logit"), data = dados)
>summary(modeloA)

## Testando a exclusão da variável Pb do modelo:
>modeloA1=glm(insucesso ~ ano + id + fm ,
family = binomial("logit"), data = dados)
>anova(modeloA1, modeloA, test="Chisq")

## Analisando a qualidade do ajustamento do modelo seleccionado:
>library(MKmisc)
>f=fitted(modeloA)
>X=cbind(Insucesso, Id, Ano, Fm, Pb)
>HLgof.test(fit=fitted(modeloA),dados$Insucesso,ngr=10,X,verbose=FALSE)
##teste de Hosmer & Lemeshow

>library(ROCR)
>predA <- prediction(f, dados$Insucesso)
>perfA<- performance(predA,"tpr", "fpr")
>plot(perfA,col="blue", xlab = "Falsos positivos",
ylab = "Verdadeiros positivos", main = "Curva de ROC- ModeloA")
##Representa a Curva ROC
>library(Hmisc)
>y=dados$Insucesso
>somers2(f, y, weights=NULL, normwt=FALSE, na.rm=TRUE) ## Calcula o valor
##AUC
>par(mfrow=c(2,2))
>plot(modeloA,wich=list(1,2,3,4,5) ##gráfico dos resíduos
```

```
## Cálculo dos odds ratio e respectivos intervalos de confiança:
>OR=exp(modeloA$coefficients)
>OR
>IC_OR=exp(confint(modeloA))
>IC_OR
```

2. Construção do Modelo B

```
>stepB=stepAIC(modelo0, direction="both",scope=list(upper=~Consumo+Id+Ano+
  Pb+Gosta+Fm+IntF+AcompF+Dific+Exig, lower0~1),trace=1)
>stepB$anova

>summary(stepB) ## para análise de significância dos coeficientes
>modB=glm(Insucesso ~ Ano + Id + Consumo + Pb + Fm,
  family = binomial("logit"), data = dados)
>summary(modB)

##Estudando interação (com o consumo)
>modB1=glm(Insucesso ~ Ano + Id + Consumo + Pb + Fm +Consumo*Id,
  family = binomial("logit"), data = dados)
>summary(modB1)

>modB2=glm(Insucesso ~ Ano + Id + Consumo + Pb + Fm +Consumo*Ano,
  family = binomial("logit"), data = dados)
>summary(modB2)
>modB3=glm(Insucesso ~ Ano + Id + Consumo + Pb + Fm +Consumo*Fm,
  family = binomial("logit"), data = dados)
>summary(modB3)

>modB4=glm(Insucesso ~ Ano + Id + Consumo + Pb + Fm +Consumo*Pb,
  family = binomial("logit"), data = dados)
>summary(modB4)

## Testando a existência de factores de confusão:
>modId=glm(Insucesso~Consumo+Id, family=binomial("logit"), data=dados)
>summary(modId)

>modAno=glm(Insucesso~Consumo+Ano, family=binomial("logit"), data=dados)
>summary(modAno)

>modFm=glm(Insucesso~Consumo+Fm, family=binomial("logit"), data=dados)
>summary(modFm)

>modPb=glm(Insucesso~Consumo+Pb, family=binomial("logit"), data=dados)
>summary(modPb)

## Aplicação do propensity score para reduzir o viés das variáveis de
confundimento (Id, Ano e Fm)

## Estimação do propensity score
>ps.logit=glm(Consumo ~ Id+Ano+Fm, family = binomial("logit"), data = dados)
>summary(ps.logit)
>X=ps.logit$fitted
>Y=dados$Insucesso

##Estratificação
>dados.s5 <- cut(X, quantile(X, seq(0, 1, 1/5)),
  include.lowest = TRUE, labels = FALSE) ## dividindo as ps.logit em 5
  grupos

##testando igualdade do valor médio dos ps.logit nos dois grupos (em cada
estrato)
>st=data.frame(Insucesso,Consumo,dados.s5)
>out=by(st,dados.s5,function(mydataframe){with(mydataframe,
  t.test(Insucesso[Consumo==0],Insucesso[Consumo==1]))
  })
>g=factor(dados.s5) ## assumindo cada estrato como uma variável dummy
```

```
## modelação do insucesso escolar em função do Consumo, assumindo a propensity
score como linear no preditor)
>modeloB=glm(Insucesso~Consumo+ Pb+g,family=binomial)
>summary(modeloB) ## (a variável PB não se mostrou significativa).

>modeloB=glm(Insucesso~Consumo+g, family=binomial) ##Modelo B Final:
>summary(modeloB)

## Análise da qualidade de ajustamento do modelo:
>G=modeloB$fitted
>M=cbind(Consumo,Id,Ano,Fm,Pb,g)
>HLgof.test(fit=fitted(modeloB),dados$Insucesso,ngr=10,M,verbose=FALSE)

>somers2(modeloB$fitted,dados$Insucesso) ## valor da curva ROC

>predB <- prediction(fitted(modeloB), dados$Insucesso)
>perfB<- performance(predB,"tpr", "fpr")
>plot(perfB,col="blue", xlab = "Falsos positivos",
      ylab = "Verdadeiros positivos", main = "Curva de ROC- ModeloB") ## Curva
      ##ROC
>par(mfrow=c(2,2))
>plot(modeloB, wish=c(1,2,3,4,5)) ## gráfico de resíduos.

##Cálculo dos odds ratio:
>OR=exp(modeloB$coefficients)
>OR
>IC_OR=exp(confint(modeloB))
>IC_OR

3. Construção do modelo C
>stepC=stepAIC(modelo0,direction="both",scope=list(
  upper=~Freq+Id+Ano+Pb+Gosta+Fm+Int+Acomp+Dific+Exg,
  lower=~1),trace=1)
>stepC$anova
>summary(stepC)

>modC=glm(Insucesso~Freq+Ano+Id+Fm+Pb, family=binomial("logit"),data=dados)
##moldeolo resultante após o teste de Wald
>summary(modC)

## Após a recategorização da variável freq e estudo da existência de
interacção:
>modeloC=glm(Insucesso~Freq+ Ano+Id +Fm+Pb+Id*Pb,
  family=binomial,data=dados)
>summary(modeloC)

## Análise da qualidade do ajustamento:

>fit=modeloC$fitted
>Y=dados$Insucesso
>X=cbind(Freq,Ano,Id,Fm,Pb)

>library(MKmisc)
>HLgof.test(fit, Y, ngr = 10, X, verbose = FALSE)
>somers2(fit,Y) ## valor da cura ROC

>predC <- prediction(fitted(modeloC), dados$Insucesso)
>perfC<- performance(predC,"tpr", "fpr")
>plot(perfC,col="blue",xlab = "Falsos positivos",
      ylab = "Verdadeiros positivos", main = "Curva de ROC- ModeloC")## curva
      ##ROC
>par(mfrow=c(2,2))
>plot(modeloC,wich=list(1,2,3,4,5)) ## gráficos de resíduos

##Cálculo de odds ratio e seus respectivos intervalos de confiança:
>OR=exp(modeloC$coefficients)
>OR
>IC=exp(confint(modeloC))
>IC
```
